

Univerzita Karlova v Praze

Přírodovědecká fakulta

Katedra biochemie



Bakalářská práce

**Příprava a charakterizace nových proteinových sekvencí
(never-born proteins)**

Expression and characterization of selected never-born proteins

Johana Šimonová

Vedoucí závěrečné práce: Mgr. Klára Hlouchová, Ph.D.

Praha, 2016

Prohlášení

Prohlašuji, že jsem tuto bakalářskou práci vypracovala samostatně pod vedením školitelky Mgr. Kláry Hlouchové, Ph.D. a všechny použité prameny jsem řádně citovala.

V Praze, dne

.....

Podpis

Poděkování

Ráda bych na prvním místě poděkovala své školitelce Mgr. Kláře Hlouchové, Ph.D. za odborné vedení, nekonečnou trpělivost, ochotu a neocenitelné rady v průběhu sepisování této práce. Děkuji za veškerý čas, který mi věnovala a za neustálou podporu.

Poděkovat bych také chtěla Doc. RNDr. Janu Konvalinkovi, CSc. za vřelý přístup a možnost vypracovat bakalářskou práci v jeho laboratoři.

Můj velký dík patří Mgr. Vjačeslavovi Tret'jačenkovi, který mě uvedl do světa práce s DNA a proteiny. Také mu děkuji za jeho čas, milý přístup a všestrannou pomoc a snahu mě vše naučit.

Další poděkování směřuji Dr. Lucii Bednárové za interpretaci CD dat, Dr. Vladimíru Kopeckému a Dr. Kateřině Hofbauerové za jejich práci při měření DLS spekter.

V neposlední řadě děkuji své rodině za podporu po celou dobu mého studia.

Abstrakt

Tato bakalářská práce se zabývá studiem náhodných polypeptidových sekvencí a výskytem sekundárních struktur v náhodném sekvenčním prostoru. Sekvence k experimentální charakterizaci byly vybrány z *in silico* knihovny obsahující 10^4 sekvencí na základě predikcí sekundárních struktur, rozpustnosti a proteinové neuspořádanosti anebo zcela náhodným výběrem.

V předkládané práci je charakterizováno 8 bílkovin ze skupiny zcela náhodného výběru sekvencí výše popsané knihovny. Nejprve je charakterizována exprese a rozpustnost těchto bílkovin. Práce se dále zaměřuje na jeden vybraný protein, u kterého byla provedena optimalizace podmínek exprese, purifikace a následná podrobnější biofyzikální charakterizace.

Klíčová slova: evoluce bílkovin, struktura bílkovin, náhodné sekvence

Abstract

This bachelor thesis deals with the study of random polypeptide sequences and occurrence of secondary structure in random sequence space. Sequences for experimental characterization were selected from library containing 10^4 random sequences on the basis of (i) either predictions of secondary structure and disorder occurrence and prediction of solubility or (ii) a completely random selection.

Eight proteins studied in this thesis belong to the second option – i.e. they were selected from the *in silico* generated library completely randomly. First, expression and solubility was analysed for these eight proteins. In addition, one of the proteins was selected for further analysis. The thesis involves optimization of its expression and purification as well as its detailed biophysical characterization.

(In Czech)

Key words: protein evolution, protein structure, random sequences

OBSAH

Abstrakt.....	4
Abstract.....	5
Seznam použitých zkratk.....	8
1 Teoretická část.....	11
1.1 Strukturní evoluce bílkovin.....	11
1.1.1 Skládání (folding) bílkovin	11
1.1.2 Neuspořádané proteiny.....	12
1.1.3 Současný svět bílkovin.....	13
1.1.4 Svět bílkovin jak ho neznáme	13
1.2 Úrovně a prvky struktury proteinů.....	14
1.2.1 Primární struktura.....	14
1.2.2 Sekundární struktura	15
1.2.2.1 α -šroubovice.....	16
1.2.2.2 β -skládaný list	17
1.2.2.3 Otočky a smyčky.....	18
1.2.3 Terciární a kvartérní struktura.....	18
1.3 Predikce struktur	19
1.4 Metody zjišťující strukturu a velikost proteinů.....	20
1.4.1 Cirkulární dichroismus (CD).....	20
1.4.2 Dynamický rozptyl světla (DLS)	21
1.4.3 Gelová permeační chromatografie (GPC).....	22
1.4.4 Hmotnostní spektrometrie	23
2 Cíle práce.....	24
3 Materiál	25
3.1 Přístroje	25
3.2 Chemikálie	25
3.3 Bakteriální kmeny	28
3.4 Pufrý a jiné roztoky	28
4 Metody	29
4.1 Práce s DNA.....	29
4.1.1 Restrikční štěpení	29
4.1.2 Horizontální agarosová elektroforéza	29

4.1.3	Izolace DNA z agarosového gelu.....	30
4.1.4	Ligace.....	30
4.1.5	Transformace kompetentních buněk.....	31
4.1.6	Izolace plasmidové DNA.....	31
4.1.7	Sekvenace DNA.....	32
4.2	Práce s proteiny.....	32
4.2.1	Expres proteinů a analýza rozpustností.....	32
4.2.2	Purifikace proteinů.....	33
4.2.3	SDS elektroforéza.....	34
4.2.4	Identifikace proteinů elektropřenosem na nitrocelulosovou membránu s následnou imunodetekcí.....	35
4.3	Příprava proteinových vzorků pro biofyzikální analýzy.....	35
4.3.1	DLS.....	36
4.3.2	CD.....	36
4.3.3	MALDI-TOF MS.....	36
4.3.4	Tryptické štěpení pro MS.....	37
5	Výsledky.....	38
5.1	Klonování 8 NBP.....	39
5.2	Rekombinantní exprese 8 NBP.....	41
5.3	Tryptické štěpení – ověření identity 8 NBP.....	42
5.4	Imunodetekce 8 NBP.....	43
5.5	Optimalizace produkce a purifikace vybraného NBP 1155.....	44
5.6	Gelová permeační chromatografie NBP 1155.....	47
5.7	CD a DLS spektrum NBP 1155.....	48
5.8	MALDI-TOF MS NBP 1155.....	50
6	Diskuze.....	51
7	Závěr.....	54
8	Reference.....	55
	Appendix.....	61

Seznam použitých zkratek

A _x	absorbance při vlnové délce x nm
angl.	anglický jazyk
APS	peroxodisíran amonný (z angl. ammonium persulfate)
BME	2-merkaptoethanol
bp	počet párů bází (z angl. base pair)
c	molární koncentrace
CBB	Coomassie Brilliant Blue
CD	cirkulární dichroismus
Da	dalton
DLS	dynamický rozptyl světla
DNA	deoxyribonukleová kyselina
DTT	dithiothreitol
<i>E. coli</i>	<i>Escherichia coli</i>
ECD	elektronová CD spektroskopie
EDTA	kyselina ethylendiamintetraoctová
FPLC	vysokorychlostní kapalinová chromatografie (z angl. Fast protein liquid chromatography)
Glu	kyselina glutamová
GPC	gelová permeační chromatografie (z angl. Gel permeation chromatography)
HCl	kyselina chlorovodíková
His	histidin

IPTG	isopropyl- β -D-thiogalaktopyranosid
kb	kilobáze, 1 kb = 1000 bp
LB medium	Luria Broth medium
Leu	leucin
LC-MS	kapalinová chromatografie s hmotnostním spektrometrem (z angl. Liquid chromatographny-mass spectrometry)
LCPL	levotočivé kruhově polarizované světlo
M	mol.dm^{-3}
MALDI	ionizace laserem za přítomnosti matrice (z angl. matrix-assisted laser desorption/ionization)
MS	hmotnostní spektrometrie (z angl. Mass spectrometry)
MW	molekulová hmotnost (z angl. Molecular Weight)
NaCl	chlorid sodný
NBP	přírodou neprobádané proteiny (z angl. Never born proteins)
NMR	nukleární magnetická resonance (z angl. Nuclear magnetic resonance)
NRMSD	střední kvadratická odchylka hodnot (z angl. normalized root mean square deviation)
OD ₅₅₀	optická denzita při vlnové délce 550 nm
PBS	fosfátový pufr (z angl. phosphate-buffered saline)
RCPL	pravotočivé kruhově polarizované světlo
RPM	otáčky za minutu (z angl. revolutions per minute)
SD	směrodatná odchylka (z angl. standard deviation)
SDS	dodecylsíran sodný (z angl. sodium dodecyle sulphate)

SDS-PAGE	elektroforéza v polyakrylamidovém gelu v prostředí SDS (z angl. sodium dodecyl sulphate polyacrylamide gel electrophoresis)
TAE pufr	tris-acetát-EDTA pufr
TCEP	tris(2-carboxyethyl)phospine
TEMED	N, N, N',N'- tetramethylethylendiamin
TOF	detektor doby letu (z angl. time of flight)
Tris	tris(hydroxymethyl)aminomethan
ÚOCHB AV ČR, v.v.i.	Ústav organické chemie a biochemie Akademie věd České republiky, v.v.i.
UK	Univerzita Karlova
UV	ultrafialové záření
λ	vlnová délka
1D	jednorozměrný
2D	dvojmrozměrný
3D	trojmrozměrný

1 Teoretická část

1.1 Strukturní evoluce bílkovin

Jedním ze zcela zásadních principů biochemie proteinů je, že *struktura bílkovin určuje funkci bílkovin*. To, že bílkoviny fungují například jako velmi výkonné katalyzátory velkého repertoáru chemických reakcí, je umožněno velmi specifickým a přesným umístěním chemických skupin do kontextu jejich trojrozměrné komplexní struktury. A to stejné umožňuje bílkovinám disponovat širokou škálou strukturních, transportních a regulačních funkcí. Protože funkce bílkovin je určena jejich strukturou, není překvapivé, že i strukturní repertoár bílkovin je velmi rozmanitý [1].

1.1.1 Skládání (folding) bílkovin

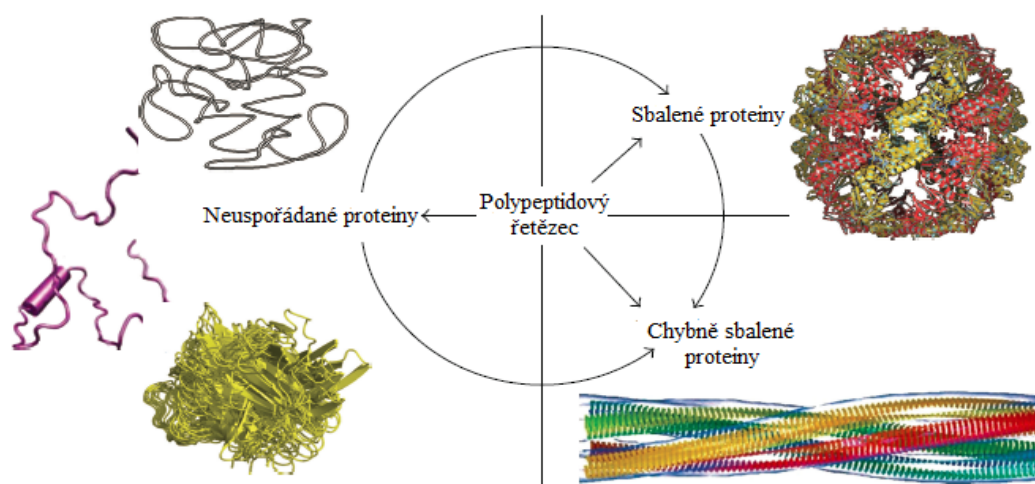
Jak je skládání proteinů do tak specifických a komplexních struktur řízeno? Tím se zabývala řada studií. Již v roce 1968 Cyrus Levinthal upozornil, že pokud by daný polypeptidový řetězec hledal nativní konformaci náhodně a prozkoumával postupně všechny možné varianty, čas potřebný k nalezení nativní konformace by byl delší než stáří vesmíru. Většina malých proteinů se ale sbalí samovolně v řádu milisekund až sekund, tento fakt je známý jako Levinthalův paradox. Levinthal soudil, že skládání bílkovin může být významně urychlováno vznikem lokálních interakcí a intermediátů [2], což bylo pro mnoho bílkovin později potvrzeno experimentálně. O několik let později Christian Anfinsen postuloval, že nativní konformaci alespoň malých bílkovin určuje pouze jejich primární struktura, tedy aminokyselinová sekvence, a sbalování bílkovin je řízeno gradientem “volné energie” [3]. Anfinsenovo dogma bývá také nazýváno termodynamická hypotéza. Na základě velkého množství teoretických i experimentálních studií v současnosti existuje několik modelů a stupňů procesu skládání bílkovin. Zahrnují počáteční fáze nukleace a vznik segmentů s uspořádanou strukturou (které na začátku značně omezí počet možných výsledných konformací), dále vznik strukturních intermediátů, domén a hydrofobního jádra a následně uspořádání podjednotek. Obecně ale stále současné teorie podporují původní princip, že finální struktura bílkoviny je určena aminokyselinovým složením a prostředím a nachází se v minimu Gibbsovy volné energie, i

když dnes například víme, že pro dosažení nativní konformace větších bílkovin je třeba asistence „šaperonů“ (z angl. chaperones) [4, 5].

1.1.2 Neuspořádané proteiny

Pro většinu bílkovin platí, že se skládají do specifické strukturní konformace, která jim slouží pro velmi specifickou funkci. Tyto proteiny nicméně často mohou zaujímat i jiné „chybně sbalené“ (z angl. misfolded) konformace buď samovolně a nebo častěji vlivem modifikací či prostředí. Tyto stavy jsou často asociovány s různými patologickými procesy [1].

Současná věda nicméně také ví, že zdaleka ne všechny biologicky aktivní bílkoviny mají ve fyziologických podmínkách nativně stabilní terciární nebo sekundární strukturu (viz Obr. 1). Výskyt těchto neuspořádaných proteinů v živých systémech je zřejmě velmi vysoký – podle současných odhadů neuspořádané proteiny tvoří 5-10% všech proteinů u bakterií a až 30% u eukaryot [6]. Funkčně tyto proteiny často komplementují strukturované proteiny, nejčastěji jejich zástupce nacházíme například v buněčné signalizaci a transkripci. Jejich strukturní flexibilita jim umožňuje interakce s různými molekulárními partnery a rovněž zaujímat různé funkce či konformace v závislosti na prostředí [6]. Potenciál široké substrátové specifity a flexibilita jsou vlastnosti, díky kterým jsou neuspořádané proteiny často uváděny jako pravděpodobní předchůdci současných bílkovin [7].



Obr. 1: Různé formy strukturního uspořádání polypeptidového řetězce. Převzato a upraveno [6].

1.1.3 Současný svět bílkovin

Evoluční stopa bílkovin se nachází v jejich sekvencích, strukturách i funkcích. Z neustále přibývajícího velkého množství počtu strukturních studií bílkovin plyne, že u vzdáleně příbuzných bílkovin se strukturní informace uchovává více než informace sekvenční [8]. Na podkladě toho byly rozčleněny evoluční vztahy mezi jednotlivými proteinovými rodinami v několika strukturních databázích. K nejznámějším patří databáze SCOP a CATH [9 - 11].

Několik milionů unikátních sekvencí bílkovin je obsaženo v neredundantních databázích sekvencí (např. Uniprot [12]). Databáze proteinových struktur (např. PDB [13]) obsahuje kolem sta tisíc vstupů. Avšak databáze SCOP a CATH klasifikují dosud poznané struktury do pouhých zhruba 1500 doménových rodin. Z toho vyplývá, že evoluce zachovala poměrně malý počet strukturních rodin i přes tak neuvěřitelnou biodiverzitu a komplexitu současných bílkovin [14]. Příčiny zachování tohoto stavu ale nejsou zcela objasněny. Je pouze pravděpodobné, že jakmile evoluce dosáhne malých a evolučně stabilních konformací, tak se ocitne pod silným selekčním tlakem tvořit stabilní a rychle se sbalující konformace. V současnosti se předpokládá, že příroda zdaleka neprobádala všechny možnosti strukturního světa bílkovin [8].

1.1.4 Svět bílkovin jak ho neznáme

Rané hypotézy předpokládají, že stávající proteiny zaujímají trojrozměrné komplexní konformace díky milionům let dlouhé evoluci a přirozenému výběru zatímco evolucí neoptimalizované sekvence budou pravděpodobně neuspořádané [15]. Podobně byla odhadována vzácnost sekvencí, které by mohli zastávat specifickou funkci (dokonce až 1 v 10^{77} sekvencích) [16, 17].

Nicméně, počet teoretických studií a experimentů z proteinového inženýrství naznačují opak [18 - 21]. Například analýza 79 sekvencí náhodně vybraných z knihovny náhodných sekvencí naznačuje, že zhruba 20 % náhodných sekvencí vykazuje vlastnosti sbalených proteinů [18]. Dokonce experimentální *in vitro* selekce náhodných sekvencí našla i hledanou funkci s frekvencí 1 sekvence v 10^{11} [22]. Je tedy zřejmé, že teorie a první studie týkající se tematiky struktury a funkce v oblasti přírodou neprozkoumaných sekvencích se mnohdy neshodují a velká část otázek zůstává nezodpovězena.

Biologické proteiny jsou v průměru tvořeny přibližně z 31 % α -šroubovicemi a 20 % β -skládanými listy [8]. Pravděpodobnost výskytu sekundárních a stabilní terciární struktury v náhodném sekvenčním prostoru je v současné době stále neznámá (ale pravděpodobně je mnohem vyšší než se předpokládalo). Ví se však, že příroda používá pro stavbu proteinů jen malý zlomek z možných sekvencí a ohromný prostor proteinových sekvencí je neprozkoumaný. Z jednoduché aritmetiky vyplývá, že pro malý protein složený ze 100 aminokyselin z biologicky kódovaných 20 aminokyselin, existuje právě 20^{100} ($\sim 10^{130}$) různých kombinací [23]. Oproti tomu se počet všech proteinů prozkoumaných přírodou od vzniku života odhaduje přibližně na 10^{21} - 10^{43} [24].

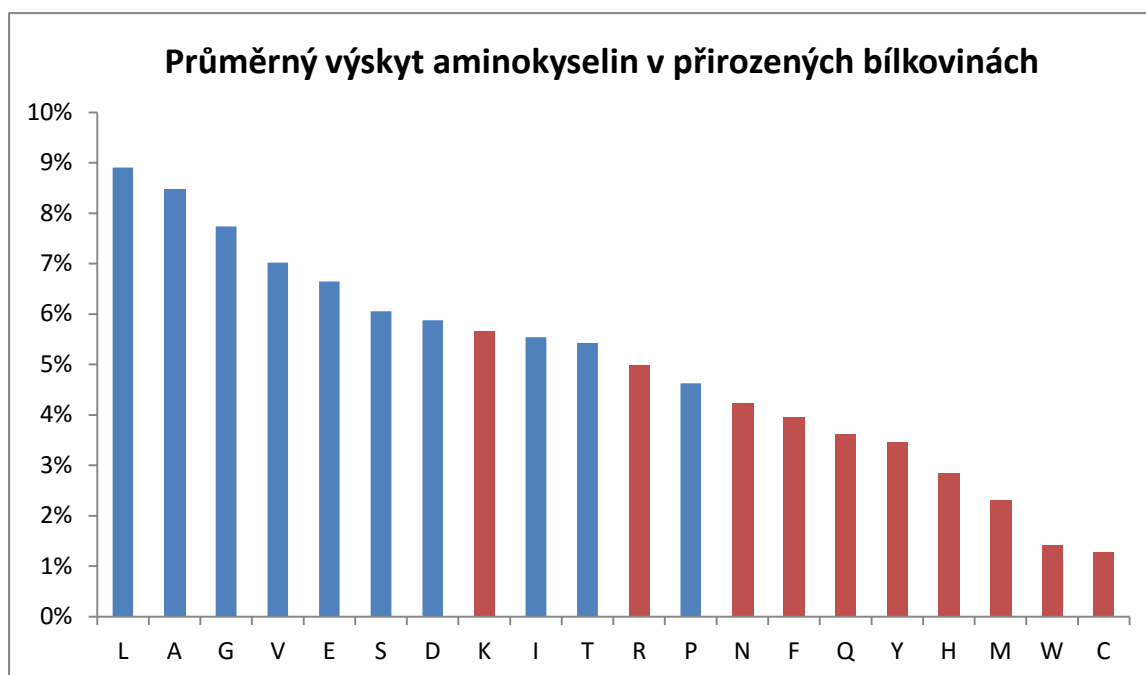
Svět nových proteinových sekvencí (z angl. never born protein, tedy NBP) a jejich strukturní potenciál je předmětem této práce. Následující kapitoly obsahují teoretický úvod k metodice, která byla použita v průběhu jejího řešení (část 1.4) anebo této práci metodicky předcházela (část 1.3).

1.2 Úrovně a prvky struktury proteinů

Současná biochemie rozlišuje až 4 úrovně struktury bílkovin, jak jsou postupně popsány níže.

1.2.1 Primární struktura

Jako primární struktura bílkovin bývá označována její sekvence aminokyselin. Právě sekvence aminokyselin je určující pro vznik konečného trojrozměrného uspořádání bílkovin. Zastoupení jednotlivých aminokyselin v přirozeně vyskytujících se bílkovinách se mírně liší mezi doménami života a v průběhu evoluce (viz Obr. 2, str. 15) [25 - 27].



Obr. 2: Průměrné zastoupení aminokyselin v proteinových sekvencích vycházejících z neredundantní databáze Uniprot [12]. Na ose x jsou uvedeny jednopísmenné zkratky aminokyselin. Modře jsou zobrazeny evolučně starší a červeně evolučně rané aminokyseliny [26, 27].

Srovnání skupiny těchto 20 aminokyselin, které dnešní život používá, s mnoha dalšími možnými skupinami aminokyselin ukázalo, že výběr těchto biologicky kódovaných aminokyselin velmi dobře pokrývá široké spektrum základních biofyzikálních vlastností (velikost, náboj a hydrofobicita) [28, 29]. Umožňuje tak dnešním proteinům prostřednictvím svých postranních řetězců disponovat širokou škálou vlastností.

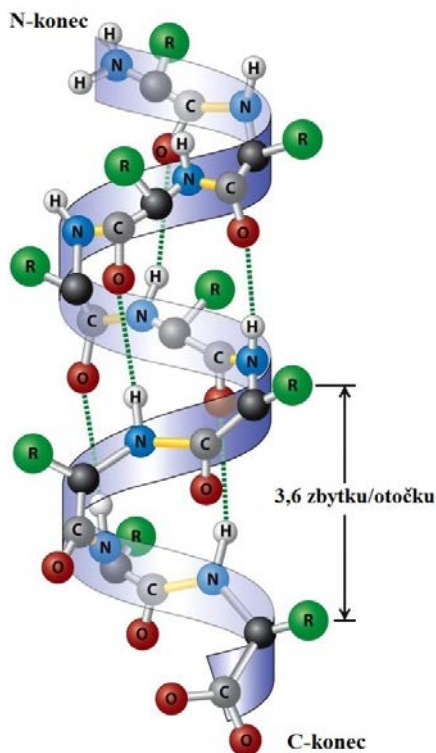
1.2.2 Sekundární struktura

Sekundární struktura, jak už název napovídá, tvoří druhý stupeň struktury proteinů. Skládání proteinů je doprovázeno tvorbou strukturně odlišných uspořádání aminokyselin. Tato uspořádání jsou omezena několika faktory. Peptidová vazba mezi dvěma zbytky není zcela pevná, s torzním úhlem (ω), který může nabývat hodnot odpovídajícím konfiguracím *cis* a *trans*. Tyto konfigurace jsou zaměnitelné, ale *trans* konfigurace je až 1000× stabilnější než *cis* (s výjimkou prolinu, který umožňuje proteinu dosáhnout jinak nemožné konfigurace) [30]. Další omezení lze najít v C_{α} -N a C_{α} -C torzních úhlech (ϕ a ψ), které mohou nabývat odlišných hodnot. Tyto hodnoty úhlů byly vypočteny G. N.

Ramachandranem z van der Waalsových poloměrů v atomech, čímž byly předpovězeny sekundární struktury. Ty byly následně empiricky potvrzeny v různých dipeptidech [31, 32].

1.2.2.1 α -šroubovice

Hlavními prvky sekundární struktury jsou smyčky, otočky, šroubovice a skládané listy. Nejběžnějším typem v proteinech je α -šroubovice. Linus Pauling první předvídal existenci α -šroubovic [33], což bylo následně potvrzeno určením první trojrozměrné struktury proteinu, myoglobinu [34]. Příklad α -šroubovice lze vidět na Obr. 3. Počet aminokyselinových zbytků připadajících na jeden závit (otočku) α -šroubovice je 3,6, což znamená, že každým 100° otočení odpovídá jeden zbytek. Výška jedné otočky měřené podél osy spirály je 0,54 nm a vzdálenost mezi ekvivalentními atomy sousedních zbytků, které tvoří hlavní řetězec, je 0,15 nm [35]. Opakující se strukturní motiv šroubovic je výsledkem opakování hodnot ϕ a ψ (-54° a -47°), které jsou pozorovány jako shlukování torzních úhlů v tzv. Ramachandrově oblasti [36]. Stabilizaci konformace šroubovice zajišťují vodíkové vazby mezi vodíkem navázaným na peptidový dusík a kyslíkem karbonylu na čtvrtém aminokyselinovém zbytku v dalším pořadí.

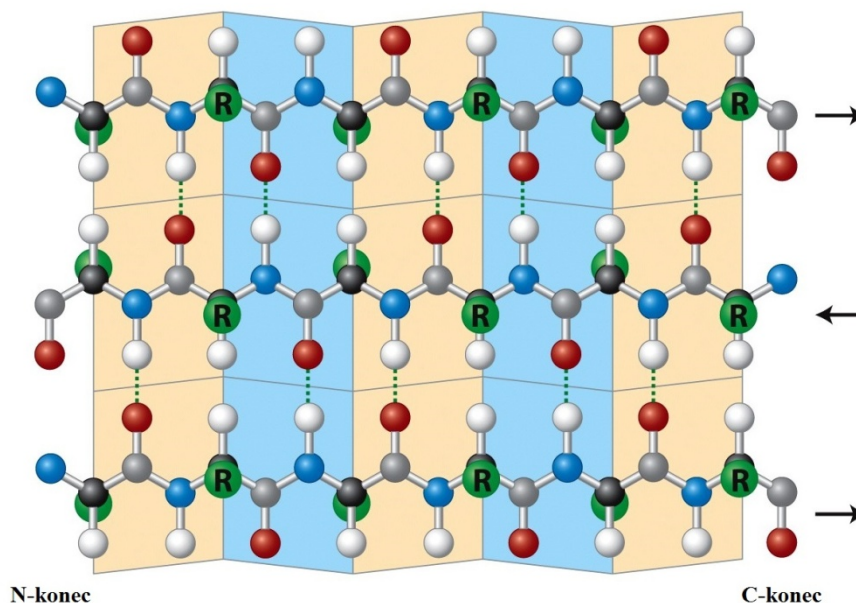


Obr. 3: α -šroubovice se znázorněným aminokyselinovým zbytkem 3,6 připadajícím na jednu otočku. Převzato a upraveno z [37].

α -šroubovice je společným sekundárním prvkem jak pro fibrilární, tak globulární proteiny. U globulárních proteinů mají šroubovice průměrné rozpětí 12 zbytků, což odpovídá více než třem spirálovým závitům. Nicméně jsou známy i šroubovice s více než 140 zbytky.

1.2.2.2 β -skládání list

Vodíkové vazby také stabilizují jiný typ sekundárních struktur – β -skládané listy. Jak je vidět z Obr. 4, vodíkové vazby spojují různé segmenty proteinové struktury. Nejsou tedy vytvořeny jen mezi sousedními zbytky, jak je tomu u α -šroubovice. Ba naopak, různé segmenty aminokyselinové sekvence, zvané β -řetězce, se shromáždí a vytvoří β -skládaný list. Lze říci, že β -skládaný list se skládá z několika β -řetězců, které drží pohromadě síť vodíkových vazeb mezi vodíkem amidu a kyslíkem karbonylu sousedních řetězců. Konformace β -skládaného listu je méně kompaktní ve srovnání s α -šroubovicí. Úhly ϕ a ψ odpovídají hodnotám -180° a 180° a vzdálenosti mezi atomy C_α jsou asi dvakrát delší než u α -šroubovice, a proto zde nedochází k intramolekulární stabilizaci [38].



Obr. 4: β -skládaný list. Převzato a upraveno z [37].

Existují dvě hlavní podoby β -skládaného listu, paralelní a antiparalelní β -skládaný list. Paralelní β -skládaný list je charakterizován dvěma sousedními peptidovými řetězci, které směřují stejným směrem. Peptidové řetězce drží pohromadě vodíkové vazby, které

jsou od sebe rovnoměrně vzdáleny, ale natačejí se mezi řetězci v šikmých úhlech. U antiparalelního β -skládaného listu směřují peptidové řetězce opačným směrem a jsou stabilizovány páry těsně nahromaděných vodíkových vazeb alternujících s páry daleko od sebe umístěných vodíkových vazeb.

β -skládaný list je označován za strukturu skládanou, protože α -uhlíky se střídavě vyskytují pod a nad rovinou listu, a proto nedochází k tvorbě makroskopických dipólových momentů jako u α -šroubovice. Dalším rozdílem oproti α -šroubovice je větší interakční vzdálenost mezi atomem dusíku z amidové skupiny ze základního skeletu a kyslíku z karbonylu. Tímto dochází ke snížení stability β -skládaného listu. Elegantním řešením, které příroda používá, je složení celého listu do větší struktury, válcovitého tvaru, nazývaného se β -barrel (β -soudek). Téměř všechny skládané listy mají zkroucené řetězce, obvykle pravotočivé. Teoretické studie uvádějí, že β -skládané listy jsou mnohem více termostabilní než α -šroubovice, a proto jsou také více odolné proti chemické degradaci [39].

1.2.2.3 Otočky a smyčky

Otočky mají uspořádanou strukturu, která se skládá ze čtyř aminokyselinových zbytků, a jsou odpovědné za změnu směru. Druhým a čtvrtým zbytkem otočky jsou obvykle *cis*-prolin a glycin. Zbylé dvě pozice otočky jsou sice méně konzervativní, ale asparagin a aspartát jsou běžnými aminokyselinovými zbytky v těchto polohách. Mezi prvním a čtvrtým zbytkem otočky se nacházejí vodíkové vazby. Smyčky jsou také odpovědné za změnu směru, ale nemají pravidelnou konformaci, je to pouze dlouhý úsek aminokyselin.

1.2.3 Terciární a kvartérní struktura

Terciární strukturu definují sekundární struktury, které jsou často organizovány do domén – kompaktních jednotek spojených polypeptidovou kostrou, neboli jsou to jednotlivé sekundární struktury uspořádané v prostoru. Podle terciární struktury se rozlišují proteiny na dva typy: globulární a fibrilární. U fibrilárních proteinů, které mají vláknitý tvar a jsou často nerozpustné, převládá právě jedna sekundární struktura. U globulárních proteinů majících převážně kulovitý tvar se vyskytují sekundární struktury jak rovnoměrně zastoupené, tak i např. všechny α , či všechny β [38].

Velká část současných bílkovin je složena z několika podjednotek, které jsou často tvořeny různými polypeptidovými řetězci. Kvartérní struktura odkazuje na prostorové uspořádání těchto podjednotek.

1.3 Predikce struktur

Jak bylo popsáno výše (Sekce 1.1.1), je veškerá informace potřebná pro konformaci proteinu obsažena v jeho primární sekvenci a prostředí. I když by to v principu mělo být možné, nedisponuje ale současná věda ještě dostatkem nástrojů pro přesnou predikci terciární struktury jen ze sekvenční informace. Současné úspěšné prediktory struktur vždy využívají kombinaci statistické teorie a empirických dat. V posledních zhruba 20 letech tato oblast prošla významným rozvojem, ale velká část tohoto vděčí rychlému nárůstu strukturních databází. Struktura, která není do značné míry podobná k již vyřešené struktuře, je v současnosti v podstatě nepredikovatelná [40, 41].

Obor bioinformatiky však v posledních letech prodělal dramatický vývoj v predikci proteinové struktury v 1D – tedy v predikci vzniku sekundární struktury. I tyto metody významně čerpají z evolučních informací obsažených ve stále se rozšiřujících databázích proteinových sekvencí a struktur. Většina prediktorů sekundární struktury pracuje na nějaké úrovni s principem, že určité segmenty po sobě jdoucích aminokyselin mají preferenci pro tvorbu určitého typu sekundární struktury. Úplně první generace prediktorů vycházela jen z propenzity jednotlivých aminokyselin pro tvorbu sekundárních struktur [42]. Další generace brala v úvahu i okolní aminokyseliny a tedy pravděpodobnost, že určitý segment (zhruba 3-50 aminokyselin) zaujme určitou sekundární strukturu. Obě tyto generace však v přesnosti predikce nepřesahovaly 60 %. Pozdější generace prediktorů začaly využívat kombinaci homologních vztahů a výkonnějších algoritmů využívajících algoritmů strojového učení (například neuronové sítě a skryté Markovovské modely) a dosahují přesnosti až 80 % [43].

1.4 Metody zjišťující strukturu a velikost proteinů

Tato práce se zabývá přípravou několika bílkovin vybraných z knihoven náhodných sekvencí (vygenerovaných *in silico*) a jejich základní biofyzikální charakterizací. V rámci této práce bylo použito jen několik z možných metod charakterizace, zejména za účelem potvrzení identity bílkovin, zjištění velikosti a oligomerního stavu a obsahu sekundárních struktur. Z toho důvodu zde nejsou podrobněji popsány metody zaměřující se například na zjištění celkové terciární struktury (rentgenová krystalografie, NMR atd.)

1.4.1 Cirkulární dichroismus (CD)

Cirkulární dichroismus patří mezi chirální spektroskopické metody. CD vychází z diferenciální absorpce levotočivého kruhově polarizovaného světla (LCPL) a pravotočivého kruhově polarizovaného světla (RCPL) s opticky aktivní látkou v závislosti na vlnové délce.

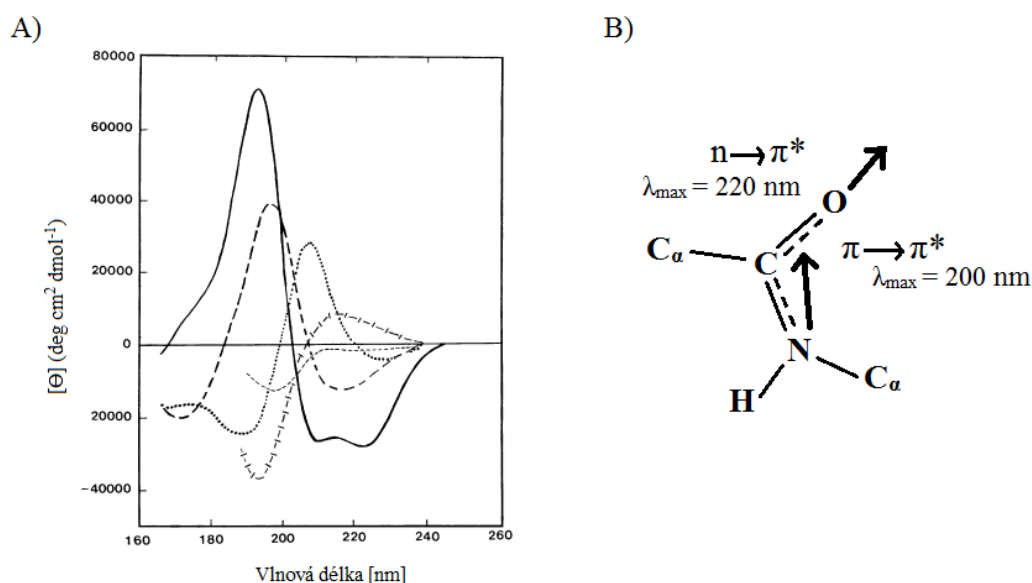
$$\Delta A = A(\lambda) \text{ LCPL} - A(\lambda) \text{ RCPL} \quad (1)$$

Proteiny obsahují řadu chromoforů, a právě proto lze při jejich studiu využívat CD spektra. Chromofory jsou látky absorbující elektromagnetické záření, častěji jen část molekuly je odpovědná za absorpci záření [44]. Některé chromofory disponují vnitřní chiralitou, jiné se nacházejí v chirálních podmínkách. CD spektra nalezneme ve vzdálené UV oblasti (240 – 280 nm), kde dochází k absorpci peptidové vazby a analyzovaná CD spektra specificky odráží vlastnosti sekundárních strukturních prvků jako je α -šroubovice a β -skládaný list. CD spektra v blízké UV oblasti (320 – 260 nm) odráží prostředí postranních řetězců aromatických aminokyselin (např. fenylalanin, tyrosin, tryptofan), které poskytují informaci o terciární struktuře proteinů.

Nejrychlejší a nejjednodušší vyhodnocení sekundární struktury poskytuje elektronová CD spektroskopie (ECD). ECD oblast charakteristickou pro určité typy sekundárních struktur lze nalézt v oblasti absorpce amidových skupin, tedy okrajová UV – pod 250 nm. Pro tuto oblast jsou typické $\pi - \pi^*$ (~190 nm až ~140 nm) a $n - \pi^*$ přechody (viz Obr. 5, str. 21). Oblast pod ~180 nm je nepřístupná pro komerční ECD nástroje. Toto omezení lze překonat synchrotronovým zdrojem záření a umožňuje ECD měření až do vakuové UV oblasti [45].

V současnosti existuje několik databází a algoritmů (např. Dichroweb [46], Protein Circular Dichroism Data Bank [47] nebo CDPro [48]), které slouží k analýze CD spekter a

jejich kvantitativnímu vyhodnocení, tedy poskytují přibližný odhad sekundárních struktur proteinu. Většina algoritmů využívá referenční databáze, které obsahují CD spektra proteinů, jejichž struktury byly vyřešeny rentgenovou krystalografií a reflektují co největší strukturní spektrum. Na základě těchto spekter algoritmy poskytují odhad zastoupení strukturních motivů ve studovaném proteinu. Výběr referenční databáze a rozsah naměřeného spektra mají často velký vliv na interpretaci zastoupení sekundárních struktur. Například dekonvoluce spekter, které jsou měřeny pouze v oblasti ≥ 200 nm neobsahují dostatek informace, aby mohl být správně interpretován obsah β -skládaných listů [49]. V daleké UV oblasti absorbuje i řada běžně používaných biologických pufrů, solí a činidel. To je často limitujícím faktorem právě při snaze získat spektrální informaci v oblasti < 200 nm. Nejsou například vhodná redukční činidla jako DTT či TCEP, protože ty také vysoce absorbují v této oblasti (190-240 nm) [50]. Při přípravě vzorků pro CD analýzu je tedy třeba věnovat zvláštní pozornost zvoleným podmínkám.



Obr. 5: A) Typická CD spektra asociována s určitými typy sekundárních struktur, B) Znázorněny jednotlivé přechody objevující se v CD spektru bílkovin. Převzato a upraveno z [45, 49].

1.4.2 Dynamický rozptyl světla (DLS)

Metoda DLS se řadí mezi fyzikální analytické metody. DLS umožňuje měřit intenzitu světla rozptýleného molekulami ve vzorku v průběhu času. Tato metoda tedy zkoumá vlastnosti a velikost částic v roztoku.

Částice se v roztoku pohybují Brownovým pohybem. Signál získaný v průběhu měření vyplývá z Rayleighova rozptylu laserového paprsku dopadajícího náhodně na částice vzorku v roztoku. V případě monodisperzního roztoku mají všechny rozptýlené vlny stejnou intenzitu, ale odlišnou fázi, díky které mezi sebou interferují. Intenzita světla dopadajícího na detektor se tedy liší podle fázového rozdílu mezi vlnami. Intenzita světla je úměrná vlnové délce světla a šesté mocnině průměru částic [51].

Intenzita rozptýleného světla se mění tak rychle, jak rychle se molekuly pohybují. Jsou známy ještě další faktory ovlivňující pohyby molekul – teplota, velikost částic a viskozita rozpouštědla. Čím vyšší je teplota a čím menší jsou částice, tím rychleji se pohybují. Čím viskóznější je rozpouštědlo, tím je pohyb částic pomalejší. Matematicky je tedy například možné ze spektra získat informaci o distribuci hydrodynamických velikostech částic v roztoku. Při měření DLS je nejdůležitější, aby vzorek nebyl kontaminovaný (např. prachovými částicemi), proto se doporučuje jeho centrifugace po dobu jedné hodiny nejméně na $15000 \times g$. Kvalita produktu závisí na koncentraci proteinu. Abychom získali spolehlivé výsledky, koncentrace vzorku musí být ideálně 1 mg/ml. Technika DLS je užitečná ke stanovení oligomerizace/agregace bílkovin, sledování proteinových interakcí či ke kontrole čistoty vzorku v průběhu krystalizace.

1.4.3 Gelová permeační chromatografie (GPC)

GPC je speciálním druhem kolonové kapalinové chromatografie, která umožňuje separovat molekuly podle jejich velikosti. GPC se nejčastěji používá ke stanovení molárních hmotností a k purifikaci směsí o různé molekulové hmotnosti.

Stacionární fázi tvoří porézní gelové částice s definovanou velikostí pórů nejčastěji na bázi polyakrylamidu, polysacharidů či zesíťovaného dextranu. Prostor mezi gelovými částicemi i póry je vyplněn rozpouštědlem tzv. mobilní fází. Rozlišujeme systémy hydrofilní a hydrofobní podle polaritý stacionární a mobilní fáze. Částice, o průměru menším než je průměr pórů, se dostanou difuzním pohybem dovnitř pórů a jsou na koloně zadržovány déle než větší částice, které jsou z kolony eluovány relativně rychle [52]. Ke stanovení správné molární hmotnosti z chromatogramu je potřeba kolonu nejdříve zkalibrovat, tedy stanovit závislost mezi elučním objemem a molární hmotností vzorků.

1.4.4 Hmotnostní spektrometrie

Základem hmotnostní spektrometrie (z angl. „mass spectrometry - MS“) je ionizace molekul a separace na základě jejich efektivní hmotnosti (m/z , kde m je hmotnost iontu a z je nábojové číslo). Tento princip je využíván v řadě aplikací, v biochemii nejčastěji ke zjištění hmotnosti biomolekul ve směsi nebo v čistém vzorku a ke stanovení jejich struktur.

Jednou z aplikací MS, která je využívána v předkládané práci, je identifikace proteinu z gelu po proteolytickém štěpení a následná identifikace získaných peptidů. Na výběr je řada různých komerčních enzymů, ale trypsin (který štěpí peptidový řetězec na C' terminálním konci argininu nebo lysinu) je nejčastěji používaným. Fragment vyříznutý z SDS-PAGE gelu je nejprve podroben redukci a alkylaci cysteinů, následně štěpen proteázou a vzniklé peptidy jsou analyzovány nejčastěji pomocí LC-MS (tedy chromatografie následovaná hmotnostní spektrometrií). Peptidová mapa je následně srovnána s databází známých peptidů a jejich hmotnostmi. Analýza peptidů pomocí MS je jednodušší než analýza proteinů, protože se lépe ionizují a jejich spektra jsou lépe interpretovatelná.

K určení molekulových hmotností proteinů se používá ionizace laserem za přítomnosti matrice (MALDI, z angl. matrix assisted laser desorption/ionization). Ve spojení s tímto způsobem ionizace se nejčastěji využívá hmotnostní analyzátor TOF (z angl. time of flight), zejména kvůli velkému rozsahu hmotností. Tato metoda je v předkládané práci použita pro stanovení a potvrzení hmotnosti čistého proteinu v roztoku [53, 54].

2 Cíle práce

- Příprava plasmidů pro rekombinantní expresi 8 proteinů vybraných z náhodné *in silico* knihovny
- Rekombinantní exprese a zjištění rozpustnosti 8 proteinů (viz výše vybraných)
- Optimalizace produkce a purifikace vybraného proteinu a jeho následná biofyzikální charakterizace metodami DLS, GPC, MS a CD

3 Materiál

3.1 Přístroje

- AKTA Explorer FPLC systém, Amersham Pharmacia Biotech, GE Healthcare Bio-Sciences (USA)
- aparatura na horizontální elektroforézu Owl, Thermo Scientific (USA)
- aparát pro vertikální elektroforézu Sigma-Aldrich (USA)
- blokový termostat MB – 102, Bioer Technology (Čína)
- CD spektrometr Jasco 815 (Japonsko)
- centrifugy: Eppendorf 5415 R (Německo)
Biofuge Pico, Heraeus Instrument (Německo)
Allegra® X-15R, Beckman Coulter (USA)
Sorvall Evolution RC, Thermo Scientific (USA)
- DLS zařízení: Zetasizer Nano ZS, Malvern Instruments Ltd. (Velká Británie)
- inkubátor IPP 400, Memmert (Německo)
- laboratorní váhy HL-400, A&D Engineering, Inc. (USA)
- pH metr, XS Instruments (Itálie)
- skenovací přístroj Typhoon 9410 Variable Mode Imager, GE Healthcare (USA)
- sonikátor Bandelin Sonoplus UW 3100 (Německo)
- spektrofotometr NanoDrop 1000, Thermo Scientific (USA)
- třepačka Innova® 44/44 R, New Brunswick Scientific (USA)
- UV transiluminátor QUANTUM ST4-1100, Vilber Lourmat (Německo)
- zdroj napětí Consort EV243, Sigma-Aldrich (USA)

3.2 Chemikálie

Bio-Rad, USA

- Biosafe Coomassie stain
- CBB G-250
- marker All Blue Precision plus protein standard
- marker Dual Xtra Standards

Biotium, Hayward, USA

- fluorescenční barvivo GelRed

Clontech, USA

- TALON

Fluka, Německo

- 2-merkaptoethanol
- Acrylamide
- Dodecylsírán sodný
- TEMED

Millipore, USA

- centrifugační filtr Ultrafree[®]-MC

New England Biolabs (NEB), Ipswich, USA

- CutSmart Buffer
- hovězí sérový albumin (BSA)
- restriční endonukleasy: NdeI, XhoI
- 10×Buffer for T4 DNA Ligase
- T4 DNA Ligasa

Penta, Prague, Česká republika

- Agarosa
- Dusičnan stříbrný
- Ethanol
- Ethylendiamintetraoctová kyselina
- Chlorovodíková kyselina
- Isopropylalkohol
- Methanol
- Sacharosa

Qiagen, Německo

- PBS Buffer
- Souprava pro purifikaci z gelu: QIAquick Gel Extraction Kit
- Souprava pro minipreparaci plasmidové DNA: QIAprep Spin Miniprep Kit
- DNA markery: GelPilot 100 bp Plus Ladder
 GelPilot 50 bp Ladder
- 5×GelPilot Loading Dye

Roth, Německo

- Bis-Tris

Sigma-Aldrich, Německo

- Dihydrogen fosforečnan sodný dihydrát
- Glycerol
- Hydrogen fosforečnan sodný dihydrát
- Chlorid sodný
- Imidazol
- LB medium
- Peroxodisíran amonný

ThermoFisher Scientific, USA

- Casein Blocker v PBS
- nanášecí pufr pro agarosovou elektroforézu: 6×DNA Loading Dye
- T7 Promoter Primer

USB, Cleveland, USA

- Glycin
- Tris

3.3 Bakteriální kmeny

- *E. coli* MACH 1 (firma Life Technologies, Praha, Česká republika)
- *E. coli* BL21 DE3 (firma Life Technologies, Praha, Česká republika)

3.4 Pufry a jiné roztoky

- TAE pufr: 57,1 ml 99% CH_3COOH , 100 ml 0,5 M EDTA, 242 g Tris-HCl, doplněno vodou na 1l, pH upraveno na 8
- pufr A: 50 mM fosfátový pufr, 30 mM NaCl, 1 mM 2-merkaptoethanol, pH 8
- pufr B: 50 mM fosfátový pufr, 30 mM NaCl, 5 mM imidazol, 1 mM 2-merkaptoethanol, pH 8
- pufr C: 50 mM fosfátový pufr, 30 mM NaCl, 250 mM imidazol, 1 mM 2-merkaptoethanol, pH 8
- pufr D: 10 mM Tris, 10 mM NaCl, 1 mM 2-merkaptoethanol, pH 8
- elektrodový pufr: 25 mM Tris-HCl, 250 mM glycín, 0,1% SDS, pH 6,8
- 6×SDS vzorkový pufr: 350 mM Tris-HCl, pH 6,8, 30% (v/v) glycerol, 350 mM SDS, 4% (v/v) 2 mM 2-merkaptoethanol, 180 μM bromfenolová modř
- 1×SDS vzorkový pufr: 6×ředěný 6×SDS vzorkový pufr
- 1×GelPilot Loading Dye: 5×ředěný 5×GelPilot Loading Dye
- Bis-Tris pufr: 1,5 mM Bis-Tris, pH 6,5-6,8 s HCl
- PBS (10×): 1,37 M NaCl, 27 mM KCl, 100 mM Na_2HPO_4 , 18 mM KH_2PO_4
- blokovací roztok: 1% (w/v) casein v 100 mM fosforečnan sodný, 150 mM NaCl, pH 7,4
- odbarvovací roztok: 10% aceton, 40% methanol doplněno destilovanou vodou do 500 ml

4 Metody

4.1 Práce s DNA

4.1.1 Restrikční štěpení

Restrikční štěpení pomocí endonukleas *NdeI* a *XhoI* bylo prováděno podle následujícího postupu. Do sterilní mikrozkušavky byla pipetována reakční směs složená z 500 ng plasmidové DNA, 1 μ l restrikční nukleasy *XhoI* (NEB, 20000 U/ml), 1 μ l restrikční endonukleasy *NdeI* (NEB, 20000 U/ml), 3 μ l komerčního pufru CutSmart (NEB) a doplněno sterilní vodou do celkového objemu 30 μ l. Reakce probíhala přes noc při 37 °C v inkubátoru. Následně byla provedena horizontální agarosová elektroforéza k izolaci a přečištění vzorků.

4.1.2 Horizontální agarosová elektroforéza

Byl připraven 2,5% (w/v) gel rozpuštěním 1 g agarosu v 40 ml 1 \times koncentrovaného TAE pufru. Tato směs byla zahřáta v mikrovlnné troubě ke snadnějšímu rozpuštění agarosu. Poté byla směs ochlazená na teplotu přibližně 50 °C a byly přidány 4 μ l fluorescenčního barviva GelRed (BIOTIUM, 10000 \times ve vodě). Směs byla promíchána, nalita na elektroforetickou desku a byl vložen elektroforetický hřeben pro vytváření příslušných jamek k nanesení vzorků. Po ztuhnutí gelu při pokojové teplotě byl hřeben vyjmut, gel byl vložen do elektroforetické vany a převrstven 1 \times koncentrovaným TAE pufrem po rysku (složení všech pufrů v Sekci 3.3).

Do první jamky bylo naneseno 5 μ l markeru GelPilot 100bp Plus Ladder (Qiagen). Do ostatních jamek byly naneseny vzorky s nanášecím pufrem 1 \times GelPilot Loading Dye (Qiagen).

Elektroforéza probíhala při napětí 120 V do té doby, než vzorky putující čelem, nedoputovaly do 75 % délky gelu. Po ukončení elektroforézy byl gel umístěn na UV transiluminátor QUANTUM ST4-1100 a pod UV světlem při 312 nm byly pozorovány DNA fragmenty.

4.1.3 Izolace DNA z agarosového gelu

DNA fragmenty byly vyříznuty z gelu skalpelem a dány do předem zvážených mikrozkuvek. Na analytických vahách byla odečtena hmotnost gelu a pomocí komerční sady QIAquick Gel Extraction Kit (Qiagen) byla prováděna izolace DNA.

Byly přidány 3 objemy „QG“ pufru na 1 objem gelu (100 mg ~ 100 µl). Vzorky byly inkubovány 10 minut při teplotě 50 °C a každé 3 minuty míchány na vortexu pro snadnější rozpuštění gelu. Po úplném rozpuštění gelu byl přidán 1 objem isopropylalkoholu a opět zamícháno. Následně byly vzorky přeneseny do sběrných mikrozkuvek s kolonkou a centrifugovány ve stolní centrifuze (60 s/13000×g). Eluovaný roztok byl odstraněn a kolonka promyta 750 µl „PE“ pufru, následovala centrifugace (60 s/13000×g). Eluovaný roztok byl opět odstraněn a kolonka opět centrifugována (60 s/13000×g), aby se odstranily zbytky roztoku „PE“. Poté byly kolonky vloženy do sterilních 1,5 ml mikrozkuvek a na střed kolonky bylo nanášeno 30 µl sterilní vody. Následovala centrifugace (2 min/13000×g) k vymytí DNA z kolonky.

4.1.4 Ligace

Ligace získaných inzertů do vektoru pET-24a byla provedena pomocí DNA ligasy z bakteriofága T4 (NEB, 400000 U/ml). Ligační směs o celkovém objemu 30 µl byla pipetována podle Tab. 1. Vektor a inzert byly smíchány v molárním poměru 1 : 3 (plasmid : inzert) k dosažení maximálního výtěžku ligace. T4 DNA ligasa byla pipetována jako poslední. Ligace probíhala v inkubátoru po dobu 12 hodin při teplotě 16 °C.

Tab. 1: Složení ligační směsi

Složka	Množství
Inzert s genem	0,06 pmol
Vektor pET24a	0,02 pmol
Pufr pro T4 DNA ligasu (10×koncentrovaný)	3 µl
T4 DNA ligasa (koncentrace 400000 U/ ml)	1,5 µl
Sterilní voda	doplněno do 30 µl

4.1.5 Transformace kompetentních buněk

Plasmidová DNA získaná ligací byla vpravena do kompetentních buněk chemickou metodou transformace. Pro účely izolace a purifikace plasmidové DNA byl použit kmen *E. coli* MACH 1, které byly připraveny na ÚOCHB AV ČR, v.v.i.. Obdobně bylo postupováno i při transformaci expresní *E. coli* linie BL-21 (DE3) za účelem produkce proteinu. Z mrazicího boxu byly vyjmuty mikrozkušavky s příslušnými kompetentními buňkami a ty byly ponechány na ledu po dobu 10 minut. Do vychlazených mikrozkušavek bylo pipetováno 40 µl kompetentních buněk a 7 µl ligační směsi. Mikrozkušavky byly ponechány na ledu po dobu 30 minut. Poté byla suspenze buněk podrobena metodě tzv. teplotního šoku (z angl. „heat shock“), kdy je cytoplasmatická membrána destabilizována působením vyšší teploty a vytvoří se v ní póry, kterými plasmid pronikne do cytosolu. Mikrozkušavky byly tedy dány do blokového termostatu nastaveného na teplotu 42 °C po dobu 90 s. Ihned poté byly vráceny na led a bylo přidáno 400 µl LB media bez antibiotika. Následovala inkubace při teplotě 37 °C 1 hodinu. Po inkubaci bylo 100 µl bakteriální kultury rozetřeno pomocí očkovací „hokejky“ na agarové plotny s kanamycinem (50 µg /ml). Plotny byly inkubovány při 37 °C po dobu přibližně 16 hodin.

4.1.6 Izolace plasmidové DNA

Plasmidová DNA byla izolována z buněk pomocí komerční sady QIAprep Spin Miniprep Kit (Qiagen). Nejdříve byla provedena kultivace v tekutém médiu, pomocí sterilní špičky byly vypíchnuty jednotlivé narostlé kolonie z agarových misek a přeneseny do 15 ml LB media s kanamycinem (50 µg/ml). Následovala inkubace v třepačce při teplotě 37 °C, 220 RPM přes noc – „noční kultury“.

Noční kultury byly odstředěny v centrifuze (15 min/3500×g), supernatant byl odstraněn, pelet resuspendován v 250 µl „Suspension“ pufru a převeden do sterilních mikrozkušavek. Bylo přidáno 250 µl „Lysis“ pufru (Qiagen), vzorky byly mírně promíchány a nechány stát 5 min při pokojové teplotě. Bylo přidáno 350 µl „Binding“ pufru (Qiagen), promícháno a ponecháno na ledu 5 min. Směs byla odstředěna ve stolní centrifuze (10 min/13000×g), supernatant přenesen do sterilní mikrozkušavky s kolonkou. Opět následovala centrifugace (60 s/13000×g), bylo přidáno 500 µl „Wash“ pufru I (Qiagen), odstředěno v centrifuze (60 s/13000×g), přidáno 700 µl „Wash“ pufru II

(Qiagen), odstředěno v centrifuze ($60\text{ s}/13000\times g$), supernatant vylit a opět centrifugace ($60\text{ s}/13000\times g$). Nakonec byla kolonka přesunuta do sterilní mikrozkušavky a plasmidová DNA byla eluována z kolonky 30 μl sterilní vody.

Koncentrace a čistota získané plasmidové DNA byla měřena při 260 a 280 nm pomocí spektrofotometru NanoDrop 1000 od firmy Thermo Scientific.

4.1.7 Sekvenace DNA

Správnost připravených vzorků byla ověřena osekvenováním provedeným firmou GATC Biotech. Na sekvenaci bylo použito 7,5 μl plasmidové DNA (40 ng/ μl) a 2,5 μl T7 promoter primeru (Thermo Scientific, 100 μM).

4.2 Práce s proteiny

4.2.1 Exprese proteinů a analýza rozpustnosti

Připravený plasmid byl transformací vnášen do bakteriální kultury *E. coli* BL-21 (DE3), které jsou určeny pro produkci proteinů, jak bylo popsáno v sekci 4.1.5.

Za účelem analytických produkcí proteinů v malém objemu bylo z nočních kultur, jejichž příprava je popsána v sekci 4.1.6, přeneseno 100 μl k 10 ml LB media s kanamycinem (50 $\mu\text{g}/\text{ml}$). Kultury byly inkubovány v třepačce při 37 °C do nárůstu kultury do dostatečné hustoty ($\text{OD}_{550} = 0,6$), přibližně po dobu 2 hodin. Poté byl přidán induktor exprese proteinu – isopropyl- β -D-1-thiaogalaktopyranosid (IPTG) tak, aby koncentrace induktoru byla 0,5 mmol/l. Produkce probíhala v třepačce při 37 °C po dobu 5 hodin. Následovalo odstředění v centrifuze ($3000\times g$, 20 min, 21 °C). V průběhu produkce byly odebrány vzorky na SDS elektroforézu, před indukcí a na konci exprese. Pelet z 1 ml kultur byl vždy resuspendován ve 120 μl 1 \times SDS vzorkového pufru.

Za účelem analýzy rozpustnosti byl buněčný pelet z celé kultury (10 ml) resuspendován v 1 ml pufru o složení 20 mM Tris, 5 mM NaCl, 1 mM TCEP, pH 8 a sonikován na ledu 3 x 10 s, intenzita 55%. Výsledný lyzát byl centrifugován ($13000\times g$, 30 min, 4 °C) a supernatant (rozpustná frakce) byl oddělen od peletu (nerozpustná frakce). Pelet byl následně resuspendován v 1 ml 1 \times SDS vzorkového pufru a pro analýzu

elektropřenosem bylo odebráno 30 μ l vzorku. Ze supernatantu bylo rovněž odebráno 30 μ l vzorku (s přídatkem 6 μ l 6 \times SDS vzorkového pufru).

V případě jednoho vybraného proteinu byla prováděna i velkoobjemová produkce, kdy bylo postupováno obdobným způsobem. Velkoobjemová produkce byla provedena v 0,2-4 l LB media s kanamycinem (50 μ g/ml) při 20 °C a s induktorem exprese proteinu IPTG o koncentraci 0,1 mmol/l (při optimalizaci těchto podmínek byly testovány i jiné podmínky, jak je popsáno v Sekci 5.4).

4.2.2 Purifikace proteinů

NBP protein 1155 byl produkován preparativně, aby mohl být dále purifikován za účelem biofyzikální charakterizace. Přesný protokol purifikace byl jedním z cílů této práce, postup byl tedy v několika aspektech různě modifikován, jak je popsáno níže.

Protein byl purifikován pouze ze solubilní frakce bakteriálního lyzátu. Ten byl připraven rozsuspendováním bakteriálního peletu v Pufru A a následnou sonikací. Sonikace probíhala na ledu v 5 cyklech (30 s, intenzita 55 %, 30 s pauza na ledu). Lyzát byl odstředěn v centrifuze (20000 \times g, 15 min, 4 °C) a supernatant, tedy solubilní frakce bakteriálního lyzátu, byl následně použit pro purifikaci proteinu. Ta byla prováděna pomocí nosiče Talon (Clontech) v chromatografickém uspořádání. Talonové kolony umožňují purifikaci proteinů s histidinovou značkou (His-tag), protože obsahuje kobalt, který váže histidiny afinitní kotvy. Obvykle byla kolona připravena nanesením 2 ml suspenze Talonu a následným promytím nosiče destilovanou vodou (2 x 5 ml) a ekvilibrací Pufrém A (2 x 5 ml). Chromatografie byla prováděna v samospádovém (z angl. “gravity flow”) uspořádání – supernatant z bakteriálního lyzátu (obvykle 10 ml) byl nanesen na kolonu, nanesený vzorek byl dále promyt 5 ml Pufru B a eluce byla provedena šestkrát 500 μ l Pufru C. Z každého kroku byl odebrán vzorek na pozdější SDS elektroforézu.

Po optimalizaci podmínek exprese byla purifikace na Talonu prováděna s gradientovou elucí (gradient 5-500 mM imidazole v Pufru C) na AKTA FPLC přístroji obdobným postupem. Pro biofyzikální charakterizaci (a rovněž analytické účely) byl vzorek dále purifikován pomocí gelové permeační chromatografie (viz Sekce 4.3).

4.2.3 SDS elektroforéza

Pro provedení SDS elektroforézy byla nejdříve sestavena aparatura a připraveny gely složením uvedeném v Tab. 2.

Tab. 2 : Složení jednotlivých gelů na SDS elektroforézu

Složka	Separací gel (12%)	Zaostřovací gel (5%)
	Objem	
Bis-Tris pufr	1,75 ml	0,9 ml
40% akrylamid (19:1)	2,1 ml	0,38 ml
Destilovaná voda	3,15 ml	1,72 ml
10% APS	70 μ l	30 μ l
TEMED	7 μ l	30 μ l

Složky 10% APS a TEMED byly přidávány až těsně před nanášením gelu. Nejprve byl napipetován mezi dvě očištěná skla umístěná v držáku skel separační roztok gelu a byl převrstven 20% ethanolem. Po 30 minutách polymerace gelu byl ethanol odlit a byl napipetován zaostřovací gel na již ztuhlý separační gel. Okamžitě byl vsunut hřeben pro vytváření jamek potřebných k nanášení vzorků. Po 30 minutách polymerace gelu byl hřeben vyjmut, skla byla umístěna do elektroforetické aparatury a převrstvena elektrodovým pufrům.

K jednotlivým vzorkům byl přidán vzorkový pufr. Takto připravené vzorky byly na 5 min povařeny při 98 °C, aby došlo k denaturaci proteinu působením SDS. Následovala centrifugace vzorků (5 min, 13000×g). Vzorky a marker byly naneseny na zaostřovací gel. Elektroforéza byla připojena ke zdroji napětí 130 V po dobu přibližně 1 hodiny. Elektroforéza byla ukončena, jakmile modrá barva doputovala ke spodnímu okraji separačního gelu.

Po ukončení elektroforézy byl gel třikrát promyt po 5 minutách v destilované vodě. Poté byl přenesen do barvicí lázně Biosafe Coomassie stain (BioRad) po dobu 60 minut. Proteiny jsou vyobrazeny v polyakrylamidovém gelu bez použití kyseliny octové nebo methanolu na fixaci a odbarvení. Vše je založeno pouze na bázi vody. Po obarvení byl gel opět přenesen do destilované vody několikrát měněné a nechal se odbarvovat přes noc.

4.2.4 Identifikace proteinů elektropřenosem na nitrocelulosovou membránu s následnou imunodetekcí

Přenos proteinů probíhal z polyakrylamidového gelu na nitrocelulosovou membránu tak, že se nejprve sestavila aparatura, tzv. “sendvič“. Filtrační papíry, membrána i gel byly promyty v blotovacím pufru a postup složení byl následovný: spodní díl soustavy (anoda), dva filtrační papíry, membrána, gel, dva filtrační papíry a horní díl soustavy (katoda). Aparatura byla naplněna blotovacím pufrem a připojena ke zdroji napětí. Samostatný přenos probíhal při 100 V po dobu jedné hodiny. Následně byla membrána blokována v blokovacím roztoku po dobu jedné hodiny. Poté byla membrána inkubována se syntetickým analogem monoklonální protilátky iBodies anti-His-tag polymerní konjugát (5 nM) v PBS pufru s přidavkem 0,1% Tween a 50×ředěný Casein Blocker (Thermo Scientific), přes noc ve 4 °C na třepačce. Základní kostrou protilátky iBodies, která byla vyrobena na ÚOCHB AV ČR, v.v.i., je kopolymer poly(N-(2-hydroxypropyl) metakrylamidu (HPMA), na který se kovalentně váže cílový ligand, afinitní kotva – biotin a zobrazovací sonda fluorofofor ATTO488 [55]. Fluorescence byla detekována při excitační vlnové délce laseru 488 nm po dobu 12 min na přístroji Typhoon 9410 variable Made Imager (GE Healthcare).

4.3 Příprava proteinových vzorků pro biofyzikální analýzy

Před biofyzikální analýzou vzorku 1155 byla provedena gelová permeační chromatografie na přístroji FPLC AKTA Explorer System (GE Healthcare). Kolona HiLoad 16/600, případně HiLoad 10/300, s nosičem Superdex 200 prep grade o objemu 120 ml (GE Healthcare) byla ekvilibrována pufrem, který byl následně použit jako mobilní fáze – 20 mM Tris, 30 mM NaCl, 1 mM TCEP, pH 8. Samostatné měření probíhalo při teplotě 23 °C při průtoku 1 ml/min. GPC bylo prováděno na ÚOCHB AV ČR, v.v.i..

Po purifikaci byl vzorek dialyzován přes noc v 1 litru Pufru D (10 mM Tris, 10 mM NaCl, 1 mM 2-merkaptoethanol, pH 8). Následně byl vzorek koncentrován ultrafiltrací přibližně na 1 mg/ml (přesná koncentrace byla zpětně určena aminokyselinovou analýzou, viz níže). V této formě byl vzorek použit pro biofyzikální analýzu na CD a DLS (viz níže)

Pro zjištění přesné koncentrace čistého proteinu byl vzorek podroben aminokyselinové analýze, kterou na ÚOCHB AV ČR, v.v.i. provedl Dr. Radko Souček. Pro vyhodnocení koncentrace bylo použito 4-5 nejvíce zastoupených aminokyselin.

4.3.1 DLS

Bezprostředně před měřením byl vzorek centrifugován při $20000\times g$ po dobu 20 minut (4 °C), aby došlo k odstranění prachových částic interferujících s měřením. Měření bylo prováděno Dr. Vladimírem Kopeckým a Dr. Kateřinou Hofbauerovou na Fyzikálním ústavu Matematicko-fyzikální fakulty, UK Praha, při konstantní teplotě 20 °C na přístroji Zetasizer Nano ZS (Malvern Instruments Ltd., Velká Británie) vybaveného He-Ne laserem o vlnové délce 633 nm. K dokonalému odstranění prachových částic byl vzorek před měřením filtrován přes 0,1 μm centrifugační filtr Ultrafree[®]-MC (Millipore, USA). Protein byl měřen v křemenné kyvetě 3 \times 3 mm o objemu 40 μl . Následně byly výsledky zpracovány na originálním softwaru Zetasizer 6.2 Malvern Instruments, Velká Británie.

4.3.2 CD

Měření CD spektra bylo provedeno pod dozorem Dr. Lucie Bednárové na ÚOCHB AV ČR, v.v.i.. CD spektra byla získána spektropolarimetrem Jasco 815 (Tokyo, Japonsko) ve spektrálním rozsahu 200 – 300 nm. Byla použita cylindrická křemenná kyveta o rozměru 0,1 cm. Měření probíhalo při pokojové teplotě s následujícím nastavením přístroje: 0,1 nm stupeň rozlišení, 5 nm/min rychlost skenování, 16 s čas odezvy a spektrální pásmo 1 nm.

Pro dekonvoluci spektra byl použit algoritmus CONTINLL na serveru CDPro (<http://sites.bmb.colostate.edu/sreeram/CDPro/>). Jako referenční databáze byl zvolen soubor SMP50 na základě vyhodnocení NRMSD parametru [48, 56].

4.3.3 MALDI-TOF MS

Stejná příprava vzorku NBP 1155 jako výše předcházela měření hmotnosti na MALDI-TOF MS v laboratoři hmotnostní spektroskopie na ÚOCHB AV ČR, v.v.i.

4.3.4 Tryptické štěpení pro MS

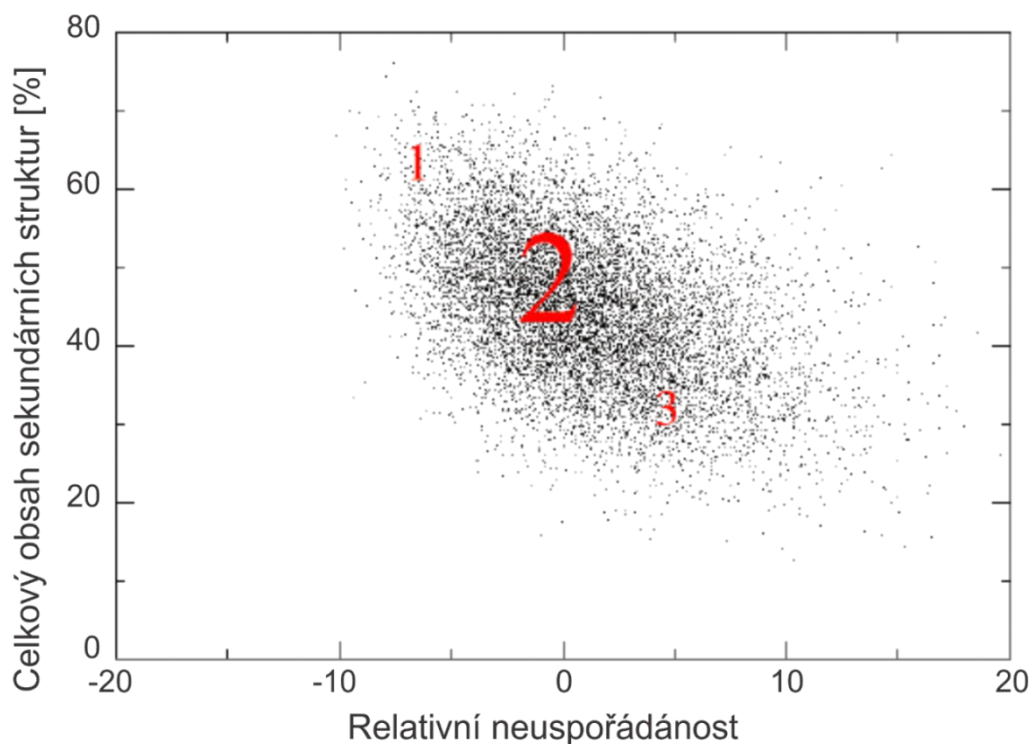
Dle Sekce 4.1.6 byly připraveny noční kultury jednotlivých NBP proteinů. Z nočních kultur bylo odebráno 400 μ l a přeneseno do 10 ml LB media s kanamycinem (50 μ g/ml). Následovala produkce dle Sekce 4.2.1. Po produkci bylo odebráno 300 μ l buněčného lyzátu a stočeno při 3000 \times g, 15 min, 21 °C. Pelet byl resuspendován ve 300 μ l 1 \times SDS vzorkovém pufru. Vzorky byly povařeny 5 min na 98 °C a nanесeny na 18% gel a elektroforéza byla připojena ke zdroji napětí 200 V na 1 hodinu. Po skončení byl gel promyt 3 \times 5 minut v destilované vodě, 30 minut barven v 0,006% (w/v) CBB G-250 (30 mg/0,5 l), 10% (v/v) kyselina octová (Bio-Rad). Poté byl gel odbarven v odbarvovacím roztoku (10% aceton, 40% methanol doplněno destilovanou vodou do 500 ml). Po odbarvení byly z gelu vyříznuty fragmenty jednotlivých NBP proteinů a předány na MS analýzu na pracovišti hmotnostní spektroskopie ÚOCHB AV ČR, v.v.i..

5 Výsledky

Předmětem této práce je příprava a experimentální analýza osmi proteinů vybraných z *in silico* knihovny náhodně vygenerovaných sekvencí. Přestože příprava a analýza této knihovny nebyla předmětem předkládané práce, je zde stručně popsána za účelem objasnění zdroje studovaných sekvencí.

Cílem celého projektu je analýza četnosti výskytu sekundárních struktur ve světě náhodných sekvencí bílkovin, tedy v oblasti „neprobádané přírodou“. *In silico* knihovny jsou tvořeny vzorkem 10^4 peptidů o délce 100 aminokyselin, kde vzájemné poměry 20 standardních aminokyselin kopírují jejich přirozené zastoupení podle databáze UniProt [16]. Jako kontrolní skupina jsou použity stejně dlouhé fragmenty strukturně charakterizovaných proteinů z databáze PDB [33]. Takto generované peptidy jsou dále postupně analyzovány několika vybranými algoritmy pro predikci sekundárních struktur (gor4, jnet, predator, simpa a psipred) a prediktory proteinové nestrukturovanosti (disopred, disembl, vsl2 a iupred). Jednotlivé knihovny jsou monitorovány z hlediska (i) četnosti sekundárních struktur, (ii) výskytu nestrukturovanosti, a (iii) poměrů jednotlivých motivů sekundárních struktur. Výsledky například ukazují, že výskyt sekundárních struktur u náhodných sekvencí je velmi blízký jejich výskytu u strukturně charakterizovaných proteinů.

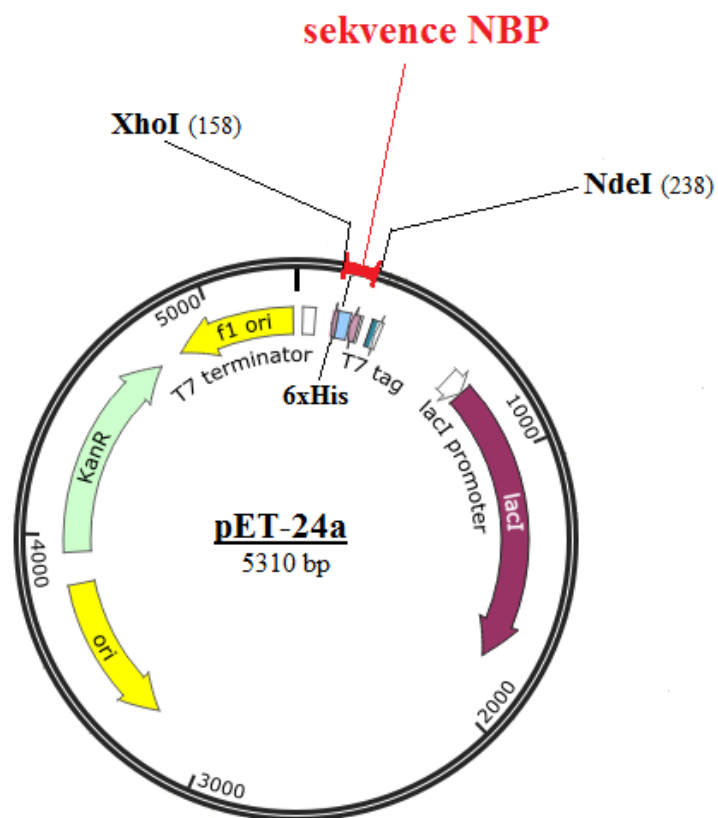
Pro účely experimentální charakterizace byly z *in silico* knihovny vybrány tři skupiny po 15 sekvencích – (1) proteiny s nejvyšším predikovaným obsahem sekundárních struktur a nejnižší nestrukturovanosti, (2) proteiny, jenž představují náhodný výběr z celé knihovny a (3) proteiny s nejnižším obsahem sekundárních struktur a nejvyšší neuspořádanosti (viz Obr. 6, str. 39). 8 NBP proteinů studovaných v této práci náleží do druhé ze zmíněných skupin.



Obr. 6: Výsledky bioinformatické predikce obsahu sekundárních struktur (osa y) a proteinové neuspořádanosti (osa x) pro knihovnu 10^4 náhodně vygenerovaných proteinových sekvencí o délce 100 aminokyselin. Na obrázku je ilustrativně vyznačen výběr 3 skupin pro experimentální charakterizaci. Autorem grafu je Dr. Jiří Vymětal (skupina Dr. Jiřího Vondráška, ÚOCHB AV ČR, v.v.i.).

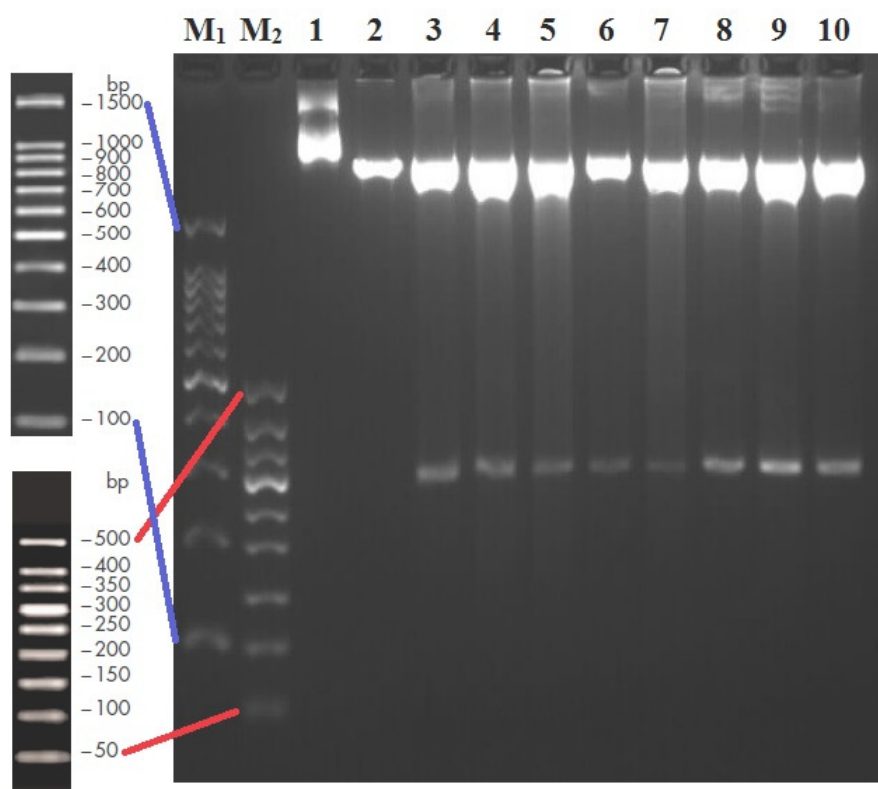
5.1 Klonování 8 NBP

Geny kódující jednotlivé NBP byly nasyntetizovány komerčně firmou Thermo Fisher Scientific (produkt GeneArt Strings DNA fragments). Geny byly zaklonovány do plasmidu pET-24a pomocí *NdeI* a *XhoI* restrikčních míst a výsledné NBP tak byly na C' konci obohaceny o aminokyseliny Leu a Glu a 6xHis kotvu (celkem tedy 108 aminokyselin) (viz Obr. 7, str. 40).



Obr.7: Plasmid pET-24a s vloženou sekvencí NBP, červeně označena sekvence NBP. Převzato a upraveno z [57].

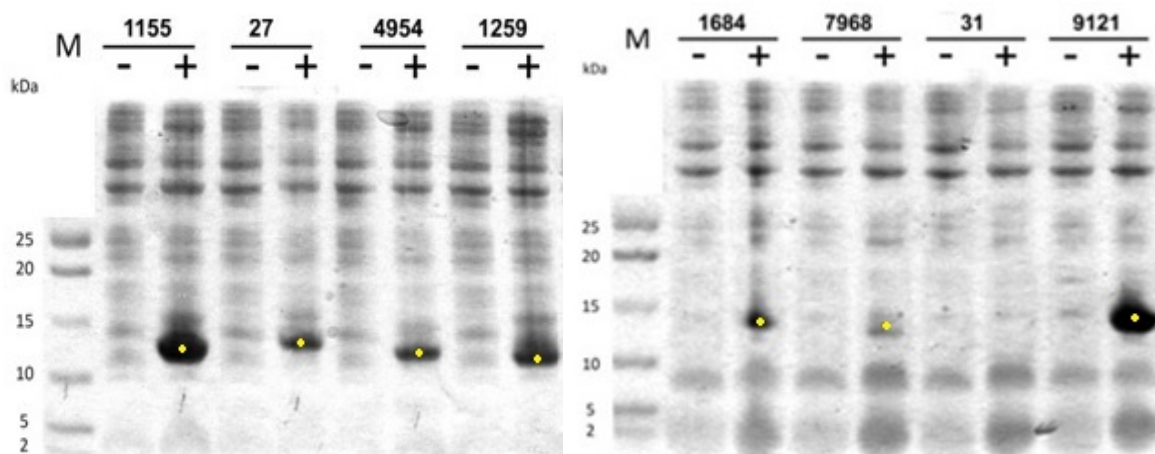
Úspěšnost klonování 8 vybraných NBP do pET-24a byla ověřena osekvenováním DNA a restrikční analýzou (viz Obr. 8, str. 41).



Obr. 8: 2,5% agarosový gel znázorňující restrikční analýzu plasmidů pET-24a-NBP; M₁ – 3 μ l markeru GelPilot 100 bp Plus Ladder; M₂ – 3 μ l markeru GelPilot 50 bp Ladder; dráha 1 – 3 μ l neštěpeného plasmidu pET-24a (170 ng/ μ l); dráha 2 – 2 μ l plasmidu pET24a naštěpeného restrikčními endonukleasami *Xho*I a *Nde*I (200 ng/ μ l); dráhy 3-10 – nanášeno 15 μ l plasmidu s geny pro proteiny 1155, 27, 4954, 1259, 1684, 7968, 31, 9121 naštěpené restrikčními endonukleasami *Xho*I a *Nde*I (zhruba 400 ng/ μ l). Obrázky markerů byly převzaty a upraveny [58].

5.2 Rekombinantní exprese 8 NBP

Expres jednotlivých NBP v buňkách *E. coli* BL-21 (DE3) byla sledována pomocí SDS-PAGE odebráním a následným porovnáním vzorků buněk před přidavkem induktoru IPTG a po kultivaci s induktorem. Analýza exprese a rozpustnosti (viz níže) byla provedena stejnou metodou pro všechny vzorky – 5 h exprese při 0,5 mM koncentraci IPTG ve 37 °C v 10 ml. V buňkách se s různou intenzitou produkovalo 7/8 testovaných NBP (viz Obr. 9, str. 42). NBP proteinům odpovídá molekulová hmotnost zhruba 12-13 kDa.



Obr. 9: SDS-PAGE analýza (v 12% redukujícím Bis-Tris gelu) produkce NBP při 37 °C, 0,5 mM koncentraci induktoru IPTG po dobu 5 hodin v 10 ml; - - 10 µl vzorku neindukované bakteriální kultury (pelet z 1 ml rozsuspendovaný v 120 µl 1×SDS vzorkového pufru); + - 10 µl vzorku indukované kultury po produkci (buněčný pelet z 1 ml rozsuspendovaný v 120 µl 1×SDS vzorkového pufru); M – 6 µl marker Dual Xtra Standards (Bio-Rad). Žlutými tečkami jsou znázorněny produkované NBP proteiny. Gely byly barveny pomocí Biosafe Coomassie stain (BioRad) (viz Sekce 4.2.3) a následně byly barevně upraveny v počítačovém programu GIMP 2 pro lepší viditelnost.

Identita všech produkujících se NBP proteinů byla ověřena hmotnostní spektroskopií (viz Sekce 5.3).

5.3 Tryptické štěpení – ověření identity 8 NBP

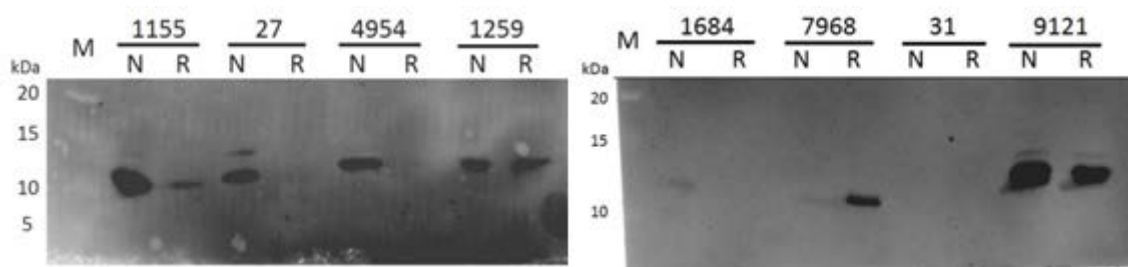
Proužky z SDS-PAGE na Obr. 9 (označeny žlutou tečkou) byly vyříznuty z gelu a po tryptickém štěpení byly podrobeny LC-MS na pracovišti ÚOCHB AV ČR, v.v.i.. Identita všech 7 exprimovaných vzorků byla takto potvrzena s 80-100% pokrytím nalezených peptidů (viz Tab. 3, str. 43).

Tab. 3: Trypsické štěpení a MS analýza studovaných NBP proteinů. Červeně je znázorněno pokrytí sekvence, které bylo nalezeno v MS spektru po trypsickém štěpení, šedě jsou znázorněny úseky sekvence, které ve spektru nebyly nalezeny.

NBP	Sekvence
27	HS AWDLVPQLRSPHLSCINMRVIKSWDLTRMVEFLDYQNTKADGFNHKMNRATEA E IATDSKARGYVFKLARTGMNSLVADLGAGKATLGHAPMKA VGLRLEHHHHHH
1155	DQTKCSSAAVS VAAHKT TNKQDADSDYFKNSTHLTPDPR VLVR NEMDTMF EVNSLI LYEHS AALTGAQISKSMKDLCTGLEFKVYPPDMIDA VQRHHVLLEHHHHHH
1259	SLGYEASED GTRVEAIKDGGRDGHMDIEAGGCGYWLFDHKDHRYAEHPMINTRDL DAQMFAMFA IFYEQDTIGFDPNPWSL ILPAVSYSMSMMEFAPAPTLEHHHHHH
1684	FWKHLTTKQAHEKETS NAMLRHAIPTA ICGIPPQFTQKTGLVLKRSPTCNMHDPA QCTRIKMFEVLRIWDVAMKMDYSGAKYYESYQCESTWGMPGAFLLEHHHHHH
4954	RVIGLSESPFDPA KFPAA SLAVGPVDGNKVPDLLQAVLALMHTSSLKNSLNIANVV RANITPRRSQPM AKRPSILFRHSSLERDLIVVWNDSIKDEALE HHHHHH
7968	VLDMPAAQKG PLQMAGTELA AEVKN GGQSTGHSLRSLRLDVAEGPSILDRLESHR ETLHDKMEAAHE LQGGKDW SLMNQYS DRGADSEHPFLKGEVTKYLE HHHHHH
9121	EDPTGVPFHHKK SHNMRKVETVNYGHKGCLVITDETNSL GVPGVDLEGLQ GAGQL AM YKPTHGIEMARA ES ELIAR KEH GKMAFIIVHRELI PEEDNRFLEHHHHHH

5.4 Imunodetekce 8 NBP

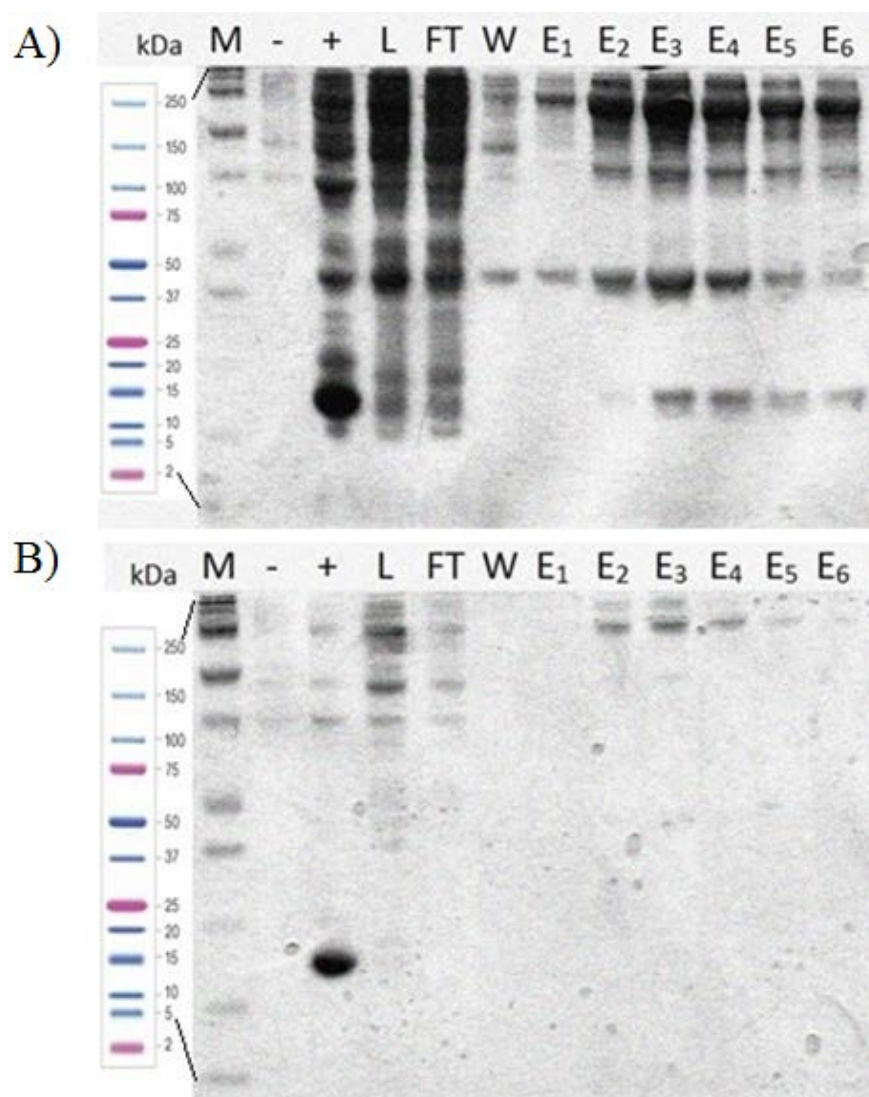
In vivo rozpustnost exprimovaných produktů byla analyzována elektropřenosem na nitrocelulosovou membránu s následnou imunodetekcí z odděleného buněčného lyzátu (na rozpustné a nerozpustné frakce po sonikaci, viz Sekce 4.2.1). Na detekci byl použit syntetický analog monoklonální protilátky iBodies anti-His-tag polymerní konjugát (25 μ M ředěna na 5 nM) s navázanou fluorescenční sondou ATTO488 (vyrobena na ÚOCHB AV ČR, v.v.i., viz Sekce 4.2.4). 4 z 8 NBP se produkují v rozpustné formě, zatímco 3 jsou pouze v nerozpustné frakci (viz Obr. 10).



Obr. 10: Imunodetekce 8 NBP na western blotu. 12% SDS-PAGE gel byl přenesen na nitrocelulosovou membránu a NBP byly detekovány pomocí iBodies anti-His-tag polymerního konjugátu s navázanou fluorescenční sondou ATTO488 (na obrázku je zobrazen výřez nízkomolekulární oblasti gelu). M – 3 μ l All Blue Precision plus protein standard (Bio-Rad); N – nerozpustná frakce buněčného lyzátu 4 μ l; R – rozpustná frakce buněčného lyzátu 6 μ l (podrobná příprava vzorků viz. Sekce 4.2.1).

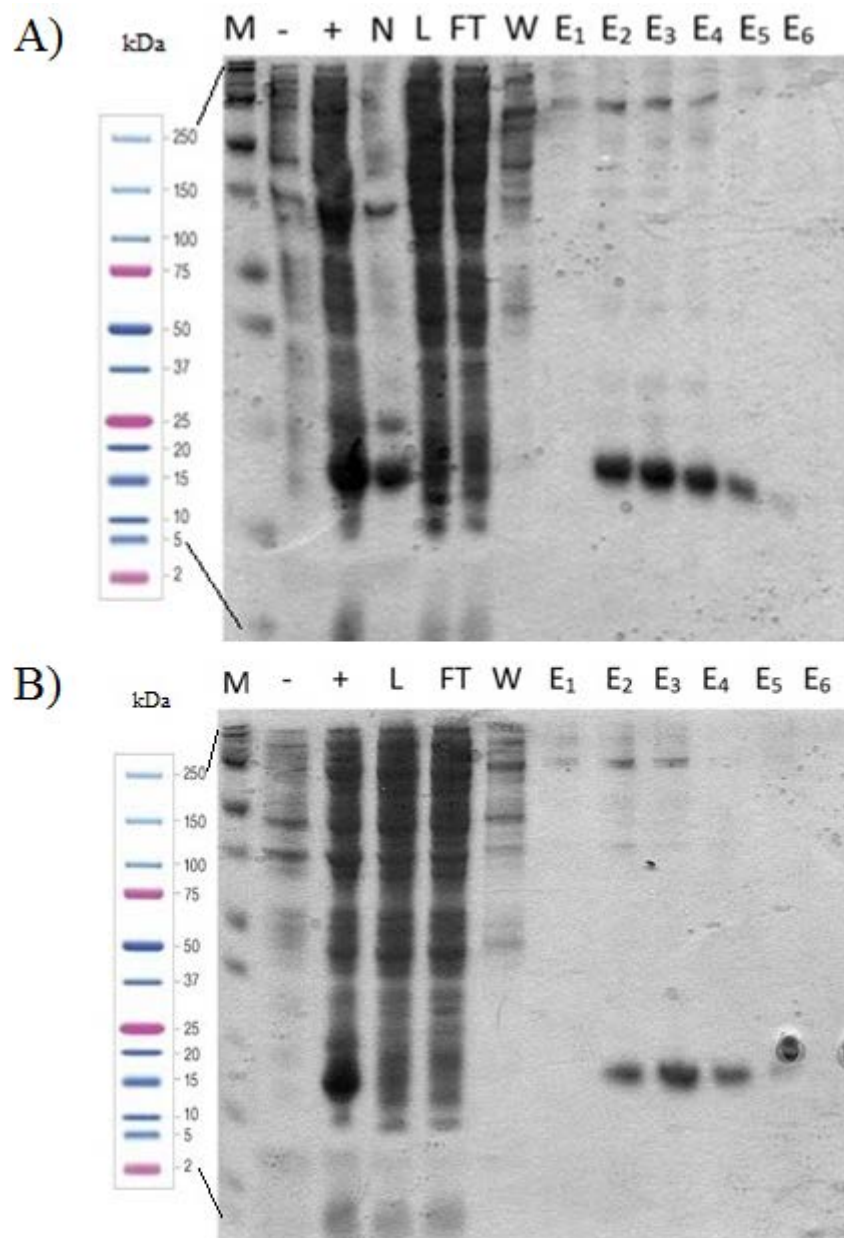
5.5 Optimalizace produkce a purifikace vybraného NBP 1155

Pro podrobnou charakterizaci byl v této práci vybrán protein 1155 pro svou silnou expresi a značně dobrou rozpustnost. Pro dosažení co největších výtěžků produkce byla exprese prováděna při různých teplotách (viz Obr. 11, str. 45) a různých koncentrací induktoru (viz Obr. 12, str. 46). Pro vyhodnocení optimalizace byl protein purifikován pouze jednokrokově za využití 6xHis kotvy na nosiči TALON v chromatografickém uspořádání (viz Sekce 4.2.2).



Obr. 11: SDS-PAGE analýza exprese a purifikace proteinu 1155, délka produkce 5 hodin při koncentraci IPTG 0,5 mM, v objemu 0,5 l a při teplotě 20 °C (A) a 37 °C (B); - - 15 μ l neindukovaná kultura, + - 8 μ l kultura po produkci (buněčný pelet z 1ml kultury byl vždy rozsuspendován a povařen ve 120 μ l 1 \times SDS vzorkového pufru); L - 8 μ l vzorek nanesen na TALON kolonu (z celkového objemu 10 ml), FT - 8 μ l, vzorek proteklý TALON kolonou (celkového objemu 10 ml), W - 10 μ l wash roztok (z celkového objemu 10 ml), E₁-E₆ - 10 μ l eluční frakce (z celkového objemu každé eluční frakce 0,5 ml).; M - 5 μ l marker Dual Xtra Standards (Bio-Rad). Obrázky markerů byly převzaty a upraveny [59].

Při stejné koncentraci induktoru, stejně dlouhé expresi, ale různých teplotách – 20 °C a 37 °C jsou výtěžky z purifikace jednoznačně vyšší při 20 °C.

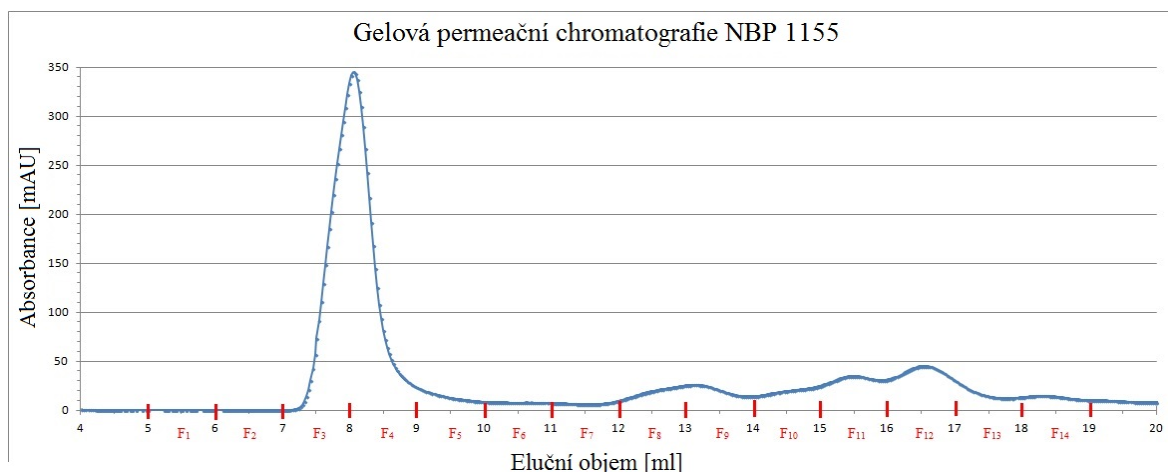


Obr. 12: SDS-PAGE analýza exprese a purifikace proteinu 1155, délka produkce 5 hodin, objem 0,2 l při teplotě 20 °C a při koncentraci induktoru IPTG 0,1 mM (C) a koncentraci 0,5 mM (D) – - 15 µl neindukovaná kultura, + - 8 µl kultura po produkci (buněčný pelet z 1ml kultury byl vždy rozsuspendován a povařen ve 120 µl 1×SDS vzorkového pufru); L – 8 µl vzorek nanesen na TALON kolonu (z celkového objemu 5 ml), FT – 8 µl, vzorek proteklý TALON kolonou (z celkového objemu 5 ml), W – 10 µl wash roztok (z celkového objemu 5 ml), E₁E₆ – 10 µl eluční frakce (z celkového objemu každé eluční frakce 0,5 ml); M – 5 µl marker Dual Xtra Standards (Bio-Rad). Obrázky markerů byly převzaty a upraveny [59].

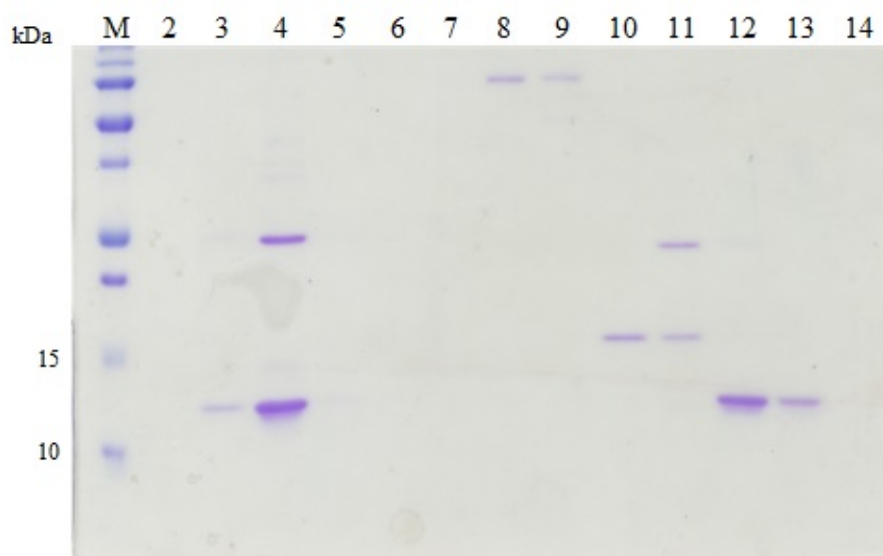
Z testovaných podmínek exprese proteinu 1155 byly pro další analýzy zvoleny následující podmínky: délka exprese 5 hodin, teplota 20 °C a koncentrace induktoru IPTG 0,1 mM. Při těchto podmínkách je výtěžek exprese a purifikace proteinu 1155 odhadována na 5 mg čistého proteinu z 1 l kultury.

5.6 Gelová permeační chromatografie NBP 1155

Za účelem přípravy proteinu pro biofyzikální charakterizaci a zjištění oligomerního zastoupení byl vzorek 1155 dále purifikován pomocí gelové permeační chromatografie (viz Sekce 4.3 a Obr. 13). Jednotlivé frakce (o objemu 1 ml) byly jímány a analyzovány pomocí SDS elektroforézy (viz Obr. 14, str. 48).



Obr. 13: Záznam průběhu gelové permeační chromatografie vzorku 1155 (kolona HiLoad 10/300, s nosičem Superdex 200 na FPLC AKTA Prima) - závislost A_{280} na elučním objemu.



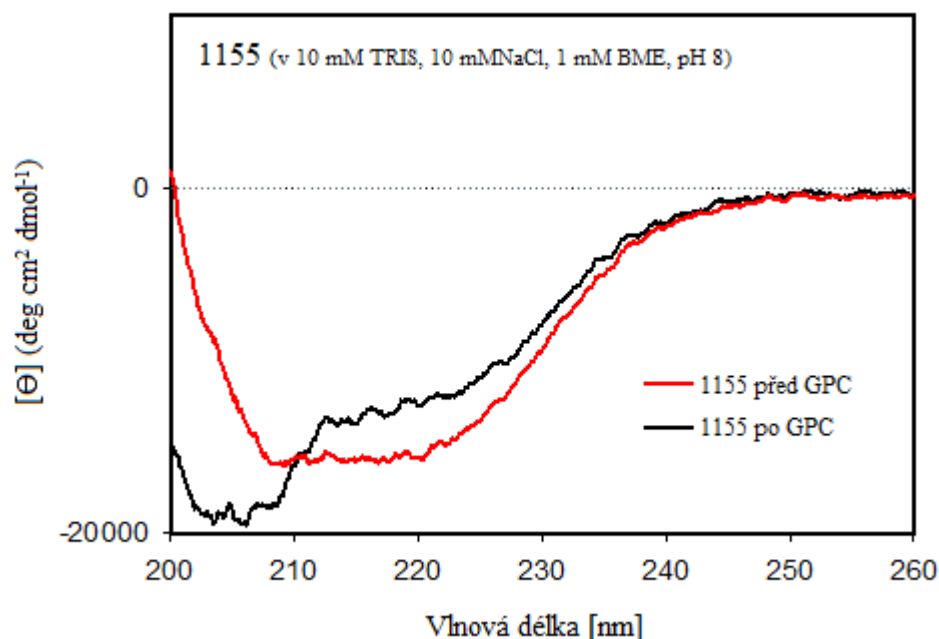
Obr. 14: SDS elektroforéza (18% redukující SDS-PAGE) zobrazující průběh gelové chromatografie. M – All Blue Precision plus protein standard (Bio-Rad), dráhy 2-14 – jednotlivé frakce F2-F14 gelové chromatografie, dráha 4 – **F₄** z Obr. 13, str. 47, dráha 12 – **F₁₂** z Obr. 13, str. 47.

Protein 1155 byl po gelové chromatografii přítomen ve dvou různých frakcích, a to při elučním objemu zhruba 8 a 16 ml. Eluční objem 8 ml odpovídá mrtvému objemu kolony a tedy se jedná zřejmě o agregovaný vzorek. Eluční objem 16 ml odpovídá podle kalibrační křivky zhruba 35 kDa.

Z toho vyplývá, že vzorek 1155 po purifikaci na nosiči Talon (tedy před gelovou chromatografií) obsahuje dvě různé oligomerní formy. Abychom byli schopni porovnat biofyzikální vlastnosti oligomerně smíšeného a nízko-oligomerního 1155, byla následná analýza (tedy CD a DLS) provedena se dvěma vzorky – “1155 před GPC” a “1155 po GPC” (frakce F₁₂ z gelové permeační chromatografie). Celkový výtěžek “1155 po GPC” je ovšem odhadován pouze zhruba na 100 µg čistého proteinu z původního 1 l nárůstu.

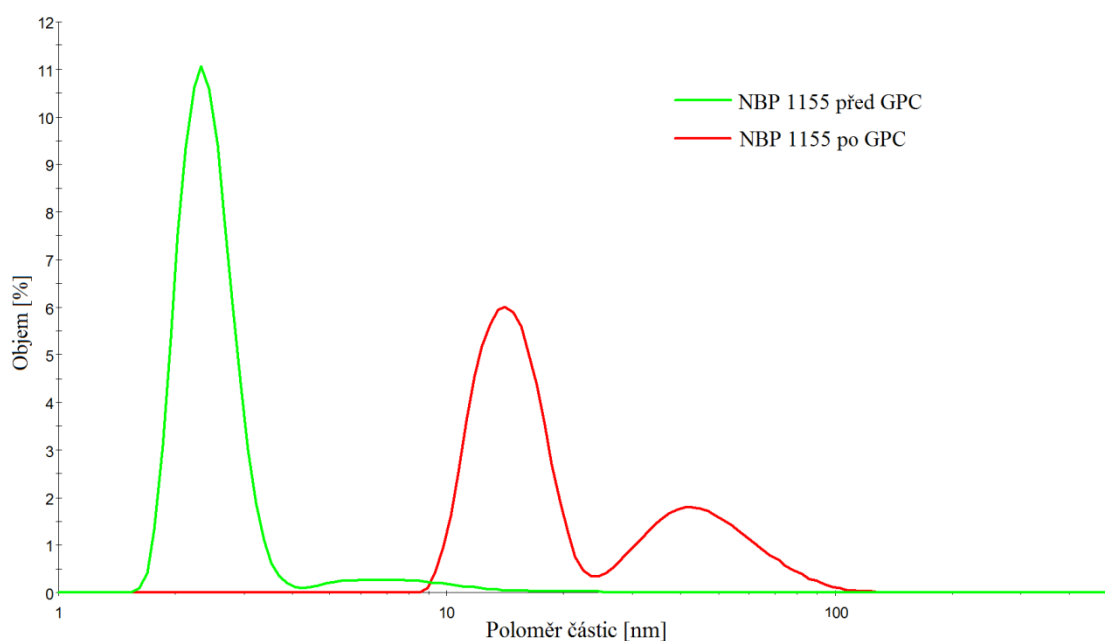
5.7 CD a DLS spektrum NBP 1155

Pro zjištění přibližného sekundárně-strukturního uspořádání vzorku 1155 (před a po gelové chromatografii) bylo změřeno CD spektrum v oblasti 200-300 nm (na Obr. 15, str. 49 zobrazeno v oblasti 200-260 nm). Na spektru je pro srovnávané vzorky přítomen zřejmý rozdíl v oblasti 200-210 nm. Vyhodnocení obsahu sekundárních struktur pomocí algoritmu CONTILL je velmi podobné pro oba vzorky (viz Tab. 4, str. 50).



Obr. 15: CD spektrální analýza NBP 1155, měření prováděno v 10 mM Tris, 10 mM NaCl, 1 mM BME a pH 8

Stejné vzorky byly použity pro zjištění velikosti přítomných částic pomocí DLS (viz Obr. 16 a také Tab. 4, str. 50). Vzorek izolovaný po gelové chromatografii odpovídá velikostí částic přibližně 1-2 oligomerům, zatímco ve vzorku před gelovou chromatografií převládají významně větší částice a vzorek byl zřejmě převážně agregovaný.



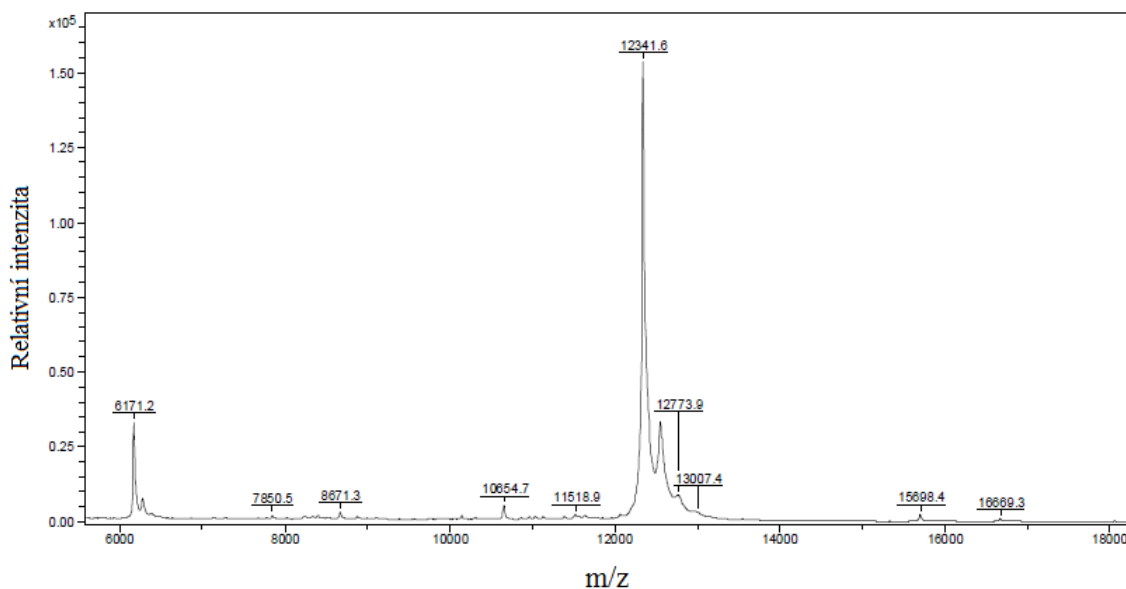
Obr. 16: DLS spektrální analýza NBP 1155, měření prováděno v 10 mM Tris, 10 mM NaCl, 1 mM BME a pH 8

Tab. 4: Přehled výsledků získaných z CD a DLS měření NBP 1155

1155	CD			DLS		
	α -šroubovice	β -skládaný list	Σ	MW \pm SD	poloměr \pm SD	Oligomer
	[%]			[kDa]	[nm]	
před gelovou chromatografií	37,4	12,0	49,4	$(23 \pm 6,4)10^2$	$16,2 \pm 3,1$	Agregát
po gelové chromatografii	37,5	12,7	50,2	30 ± 16	$2,6 \pm 0,4$	mono-/dimer

5.8 MALDI-TOF MS NBP 1155

Střední hodnota hmotnosti proteinu NBP 1155 (monoizotopická hmotnost 12335 Da) byla stanovena pomocí MALDI-TOF hmotnostní spektroskopie na 12342 Da (viz Obr. 17).



Obr. 17: MALDI-TOF MS spektrum NBP 1155

6 Diskuze

Hlavním cílem předkládané práce bylo analyzovat expresi a rozpustnost osmi proteinů, které byly vybrány z knihovny náhodných polypeptidových sekvencí. Tyto knihovny byly před zahájením této práce generovány *in silico* za účelem zjištění výskytu sekundárních struktur v náhodném sekvenčním prostoru. Celá knihovna obsahovala 10^4 sekvencí o délce 100 aminokyselin a vzájemné poměry jednotlivých aminokyselin odpovídaly výskytu aminokyselin v biologických proteinech. Celá knihovna byla ve spolupráci se skupinou Dr. Jiřího Vondráška (ÚOCHB AV ČR, v.v.i.) charakterizována prediktory sekundárních struktur a také prediktory neuspořádanosti. Na základě těchto dat byly vybrány 3 skupiny (každá obsahuje 15 sekvencí) proteinů: (1) proteiny s nejvyšším obsahem sekundárních struktur a nejnižší neuspořádaností, (2) zcela náhodný výběr a (3) proteiny s nejnižším obsahem sekundárních struktur a nejvyšší neuspořádaností. Cílem celého projektu bylo experimentálně produkovat všechny tyto skupiny za účelem (i) zjištění exprimovatelnosti/rozpustnosti, (ii) biofyzikální charakterizace a zjištění obsahu sekundárních struktur v porovnání s predikcemi a (iii) výběrem několika proteinů pro strukturní studie. V předkládané práci bylo charakterizováno 8 proteinů z druhé experimentální skupiny.

Geny jednotlivých proteinů (syntetizovány komerčně tak, že kodony byly optimalizovány pro expresi v *E. coli*) byly úspěšně zaklonovány do plasmidu pET-24a a exprimovány s His-kotvou na C' konci. His-kotva byla vybrána na základě malé velikosti a ceny souvisejících produktů (například protilátka a purifikační matrice), aby bylo možné postup purifikace použít pro velké množství vzorků. Abychom mohli studované sekvence (které měly v experimentu na N' konci navíc Met a na C' konci po klonování aminokyseliny Leu, Glu a 6xHis) srovnávat s predikcemi pro původní *in silico* sekvence, byla predikce zopakována pro tyto upravené sekvence a bylo potvrzeno, že tyto aminokyseliny neovlivňují předpovězený obsah strukturních motivů.

Proteiny byly produkovány v *E. coli* kmenu BL21 DE3. V alespoň malém množství se produkovalo 7/8 proteinů a z toho 4 byly v solubilní formě. Jedním z cílů bylo porovnání exprese a rozpustnosti proteinů mezi jednotlivými skupinami. Na základě výsledků mých spolupracovníků jsme mohli porovnat, že nejvíce bílkovin a s nejvyšší rozpustností se produkuje v první skupině a nejméně ve třetí skupině, zatímco druhá

skupina vykazuje vlastnosti přibližně průměrné. To bylo poměrně očekávatelné vzhledem k tomu, že bílkoviny s nízkým obsahem struktury jsou uvnitř buňky náchylnější degradaci.

Jeden z proteinů (NBP 1155) byl vybrán na podrobnější biofyzikální analýzu. Celkově se produkoval velmi dobře - po optimalizaci podmínek exprese byl z jednoho litru nárůstu výtěžek zhruba 5 mg čistého proteinu. Gelová permeační chromatografie a následná DLS analýza ale ukázaly, že velká část NBP 1155 je zřejmě zagregovaná a pouze malá část (zhruba 100 μ g z 1 litru nárůstu) je nízkomolekulární, tedy monomer či dimer. Toto nelze přesně určit kvůli nízké molekulové hmotnosti proteinu (zhruba 12 kDa) a velké chybě stanovení jak při GPC tak DLS.

Před měřením CD spektra NBP 1155 byly tyto dvě formy odděleny preparativně a následná analýza (CD i DLS) byla provedena se vzorkem před GPC a po GPC (tedy s nízkomolekulární formou). Výsledky z CD analýzy vzorku před GPC a po GPC se lišily jen zanedbatelně, i přes zjevné rozdíly v jejich spektru v oblasti okolo 200 nm. V obou případech bylo algoritmem CONTINLL vyhodnoceno zhruba 37 % α -šroubovice a 12 % β -skládaného listu. To by znamenalo, že agregace NBP 1155 nemá významný vliv na tvorbu sekundární struktury. Podle bioinformatické analýzy obsahuje NBP 1155 35 % α -šroubovice, 15 % β -skládaného listu, celkem tedy 50 % sekundárně strukturních motivů. Bioinformatická a CD analýza se ve výsledcích tedy s malou nepřesností shodují, nicméně tato data by měla být interpretována s velkou opatrností. Bioinformatická predikce vybranými algoritmy má přesnost zhruba 80%. Analýza CD spektra měřeného pouze do 200 nm má rovněž velmi omezenou schopnost kvantifikovat obsah strukturních motivů, zejména β -skládaného listu [49]. CD spektrum prezentované v této práci bylo součástí předběžné charakterizace (z toho důvodu není ani uváděna směrodatná odchylka) a v současné době jsou spektra přeměřována podrobněji až do oblasti 185 nm. Nicméně v celkové srovnávací studii 3 studovaných skupin není prováděna kvantitativní analýza jednotlivých spekter a spektra jsou vyhodnocována spíše kvalitativně. CD spektrum NBP 1155 tedy potvrdilo, že se jedná o protein s poměrně velkým obsahem sekundární struktury, zejména helikální.

1D NMR spektrum NBP 1155 (měřeno na ÚOCHB AV ČR, v.v.i.) navíc vykazuje náznak hydrofóbního jádra (tato data nejsou v předkládané práci uváděna – jednalo se o předběžné měření) a v současné době se snažíme připravit protein ve větším množství, aby mohla být struktura proteinu podrobněji zkoumána pomocí 2D/3D NMR. Je možné zkoušet další optimalizaci exprese (například v jiné buněčné linii nebo v přítomnosti

činidel, které mohou ovlivnit stupeň agregace), abychom snížili výskyt agregace a zvýšili výtěžek mono-/dimerní formy. Prokáže-li podrobnější měření NMR spektra přítomnost hydrofobního jádra a terciární uspořádání, je NBP 1155 jedním z prvních nahlédnutí do struktur nebiologických proteinů a může být oknem do světa přírodou neprobádaných uspořádání.

7 Závěr

Osm sekvencí DNA kódujících proteiny vybraných z náhodné *in silico* knihovny bylo úspěšně zaklonováno do plasmidu pET-24a a příslušné proteiny byly produkovány v *E. coli*. 7/8 proteinů se exprimovalo, z toho 2 ve velkém množství. Identita jednotlivých proteinů byla potvrzena hmotnostní spektroskopií.

Byly nalezeny optimální podmínky produkce a purifikace vybraného NBP 1155. Největší výtěžky proteinu 1155 byly nalezeny při podmínkách exprese: exprese 5 hodin, teplota 20 °C, koncentrace induktoru IPTG 0,1 mM.

Byla provedena biofyzikální charakterizace NBP 1155 metodami GPC, MS MALDI, CD a DLS. Metodami GPC a DLS bylo zjištěno, že majoritní část proteinu je zagregovaná a preparativně byla oddělena nízkomolekulární frakce (monomer či dimer). CD analýzou bylo zjištěno, že protein NBP 1155 obsahuje 49,8% sekundárně strukturních motivů, z toho 37,5% α -šroubovice a 12,4% β -skládaného listu. Tyto hodnoty jsou velmi blízké výsledkům bioinformatické analýzy.

8 Reference

- [1] Berg, J., Tymoczko, J. & Stryer, L. *Biochemistry*, 7th edition. W.H. Freeman and Company, chapter 2 (2007).
- [2] Levinthal, C. Are There Pathways for Protein Folding? *Extr. du J. Chim. Phys.* **65**, 44 (1968).
- [3] Anfinsen, C. B. Principles that Govern the Folding of Protein Chains. *Science* **181**, 223–230 (1973).
- [4] Ellis, R.J., Dobson, Ch., Hartl, U. Sequence does specify protein conformation. *Trends in Biochemical Sciences* **23**, 468 (1998).
- [5] Dobson, C. M. & Karplus, M. The fundamentals of protein folding: bringing together theory and experiment. *Elsevier Scienc*, 92–101 (1999).
- [6] Uversky, V. N. The mysterious unfoldome: Structureless, underappreciated, yet vital part of any given proteome. *Journal of Biomedicine and Biotechnology* 2010, 568068; doi: 10.1155/2010/568068 (2010).
- [7] James, L. C. & Tawfik, D. S. Conformational diversity and protein evolution - A 60-year-old hypothesis revisited. *Trends in Biochemical Sciences* **28**, 361–368 (2003).
- [8] Gu, J., Bourne, P. *Structural Bioinformatics*. 2th edition. Wiley-Blackwell, chapter 17, 475 (2009).
- [9] Levitt, M. Nature of the protein universe. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 11079–11084 (2009).
- [10] Hubbard, T. J. P., Ailey, B., Brenner, S. E., Murzin, A. G. & Chothia, C. SCOP: A structural classification of proteins database. *Nucleic Acids Res.* **27**, 254–256 (1999).

- [11] Sillitoe, I., Lewis, T. E., Cuff, A., Das, S., Ashford, P., Dawson, N. L., Furnham, N., Laskowski, R. A., Lee, D., Lees, J. G., Lehtinen, S., Studer, R. A., Thornton, J., Orengo, Ch. A. CATH: Comprehensive structural and functional annotations for genome sequences. *Nucleic Acids Res.* **43**, D376–D381 (2015).
- [12] Uniprot, <http://www.uniprot.org/>.
- [13] Protein Data Bank, <http://www.rcsb.org/pdb/home/home.do>.
- [14] Orengo, C. A. & Thornton, J. M. Protein families and their evolution – A structural perspective. *Annu. Rev. Biochem.* **74**, 867–900 (2005).
- [15] Meier, S. & Ozbek, S. A biological cosmos of parallel universes: Does protein structural plasticity facilitate evolution? *BioEssays* **29**, 1095–1104 (2007).
- [16] Axe, D. D. Extreme functional sensitivity to conservative amino acid changes on enzyme exteriors. *J. Mol. Biol.* **301**, 585–595 (2000).
- [17] Axe, D. D. Estimating the prevalence of protein sequences adopting functional enzyme folds. *J. Mol. Biol.* **341**, 1295–1315 (2004).
- [18] De Lucrezia, D., Franchi, M., Chiarabelli, C., Gallori, E. & Luisi, P. L. Investigation of de novo totally random biosequences. Part IV. Folding properties of de novo, totally random RNAs. *Chem. Biodivers.* **3**, 869–877 (2006).
- [19] Minervini, G., Evangelista, G., Villanova, L., Slanzi, D., De Lucrezia, D., Poli, I., Luisi, P. L., Polticelli, F. Massive non-natural proteins structure prediction using grid technologies. *BMC Bioinformatics* **10 Suppl 6**, S22 (2009).
- [20] LaBean, T. H., Butt, T. R., Kauffman, S. A. & Schultes, E. A. Protein folding absent selection. *Genes (Basel)*. **2**, 608–626 (2011).

- [21] Chiarabelli, C., De Lucrezia, D., Stano, P., Luisi, P. L. The World of the “Never Born Proteins“. *Origins of Life and Evolution of Biospheres* **39**, 308–309 (2009).
- [22] Keefe, A. D. & Szostak, J. W. Functional proteins from a random-sequence library. *Nature* **410**, 715–718 (2001).
- [23] Luisi, P. L. *THE EMERGENCE OF LIFE: From Chemical Origins to Synthetic Biology*. Cambridge University Press, 46 (2007).
- [24] Dryden, D. T. F., Thomson, A. R. & White, J. H. How much of protein sequence space has been explored by life on Earth? *J. R. Soc. Interface* **5**, 953–956 (2008).
- [25] Rost, B. Did evolution leap to create the protein universe? *Curr. Opin. Struct. Biol.* **12**, 409–416 (2002).
- [26] Trifonov, E. N. Consensus temporal order of amino acids and evolution of the triplet code. *Gene* **261**, 139–151 (2000).
- [27] Higgs, P. G. & Pudritz, R. E. A thermodynamic basis for prebiotic amino acid synthesis and the nature of the first genetic code. *Astrobiology* **9**, 483–490 (2009).
- [28] Philip, G. K. & Freeland, S. J. Did evolution select a nonrandom ‘alphabet’ of amino acids? *Astrobiology* **11**, 235–240 (2011).
- [29] Freeland, S. ‘Terrestrial’ Amino Acids and their Evolution. *Amino Acids, Peptides and Proteins in Organic Chemistry* **1**, 43–75 (2010).
- [30] Andreotti, AH. Native state proline isomerization: An intrinsic molecular switch. *Biochemistry* **42**, 9515-9524 (2003).
- [31] Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* **7**, 95–99 (1963).

- [32] Ramachandran, G. N. & Sasisekharan, V. Conformation of polypeptides and proteins. *Adv. Protein Chem.* **23**, 283–438 (1968).
- [33] Pauling, L., Corey, R. B. & Branson, H. R. The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain. *Proc. Natl. Acad. Sci. U. S. A.* **37**, 205–211 (1951).
- [34] Strandberg, B. Chapter 1: Building the Ground for the First Two Protein Structures: Myoglobin and Haemoglobin. *J. Mol. Biol.* **392**, 2–10 (2009).
- [35] Murray, R. K., Bender, D. A. & Weil, A. P. *Harper's Illustrated Biochemistry*, 28th edition. The McGraw-Hill Companies (2009).
- [36] Crieghton, T. Proteins: Structures and molecular properties. *New York W. H. Free. Co.* **10**, 82 (1992).
- [37] Lodish, H., Berk, Zipursky, A. L., Matsudaira, P., Baltimore, D., J. E. D. *Molecular Cell Biology*. 6th edition. W.H. Freeman and Company (2007).
- [38] Voet, D. & Voet J. *Biochemistry*, 4th edition. Wiley-Blackwell (2011).
- [39] Brack, A. & Orgel, L. E. Beta structures of alternating polypeptides and their possible prebiotic significance. *Nature* **256**, 383–387 (1975).
- [40] Moult, J., Fidelis, K., Rost, B., Hubbard, T. & Tramontano, A. Critical assessment of methods of protein structure prediction (CASP)--round 6. *Proteins* **61 Suppl 7**, 3–7 (2005).
- [41] Fersht, A. *Structure and Mechanism in Protein Science: a guide to enzyme catalysis and protein folding*. 2nd edition. H. Freeman and Company (1999).
- [42] Chou, P. Y. & Fasman, G. D. Prediction of protein conformation. *Biochemistry* **13**, 222–245 (1974).

- [43] Gu, J., Bourne, P. *Structural Bioinformatics*. 2nd edition. Wiley-Blackwell, chapter 29, 679–684 (2009).
- [44] Kodíček, M. & Karpenko, V. *Biofysikální chemie*. 2. vydání. Academia (2000).
- [45] Kelly, S. M., Jess, T. J. & Price, N. C. How to study proteins by circular dichroism. *Biochim. Biophys. Acta - Proteins Proteomics* **1751**, 119–139 (2005).
- [46] Whitmore, L. & Wallace, B. A. DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res.* **32**, 668–673 (2004).
- [47] Whitmore, L., Woollett, B., Miles, A. J., Klose, D. P., Janes, R. W., Wallace, B. A. PCDDb: The protein circular dichroism data bank, a repository for circular dichroism spectral and metadata. *Nucleic Acids Res.* **39**, 480–486 (2011).
- [48] Sreerama, N. & Woody, R. W. Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal. Biochem.* **287**, 252–260 (2000).
- [49] Whitmore, L. & Wallace, B. A. Protein secondary structure analyses from circular dichroism spectroscopy: Methods and reference databases. *Biopolymers* **89**, 392–400 (2008).
- [50] Jr, W. C. J. Protein secondary structure and circular dichroism: a practical guide. *Proteins Struct. Funct. Genet.* **7**, 205–214 (1990).
- [51] Some, D. Light-scattering-based analysis of biomolecular interactions. *Biophys. Rev.* **5**, 147–158 (2013)
- [52] Churáček, J. *Analytická separace látek*. 1. vydání. Státní nakladatelství technické literatury (1990).
- [53] Caprioli, R., Farmer, T. and Gile, J. Molecular Imaging of Biological Samples: Localization of Peptides and Proteins Using MALDI-TOF MS. *Anal. Chem.*, **69**, 4751–4760 (1997).

- [54] Medzihradszky, K. F., Campbell, J. M., Baldwin, M. A., Falick, A. M., Juhasz, P., Vestal, M. L., Burlingame, A. L. The characteristics of peptide collision-induced dissociation using a high-performance MALDI-TOF/TOF tandem mass spectrometer. *Anal Chem* **72**, 552–558 (2000).
- [55] Šácha, P., Knedlík, T., Schimer, J., Tykvart, J., Parolek, J., Navrátil, V., Dvořáková, P., Sedlák, F., Ulbrich, K., Strohalm, J., Majer, P., Šubr, V., Konvalinka, J. IBodies: Modular synthetic antibody mimetics based on hydrophilic polymers decorated with functional moieties. *Angew. Chemie - Int. Ed.* **55**, 2356– 2360 (2016).
- [56] Johnson, W. C. Analyzing protein circular dichroism spectra for accurate secondary structures. *Proteins Struct. Funct. Genet.* **35**, 307–312 (1999).
- [57] [https://www.snapgene.com/resources/plasmid_files/pet_and_duet_vectors_\(novagen\)/pET-24a\(+\).](https://www.snapgene.com/resources/plasmid_files/pet_and_duet_vectors_(novagen)/pET-24a(+).) (2.7.2016).
- [58] [https://www.qiagen.com/us/search/gelpilot-dna-molecular-weight-markers/.](https://www.qiagen.com/us/search/gelpilot-dna-molecular-weight-markers/) (24.6.2016).
- [59] [http://www.bio-rad.com/en-cz/product/prestained-protein-standards.](http://www.bio-rad.com/en-cz/product/prestained-protein-standards) (24.6.2016).

Appendix

Sekvence studovaných bílkovin

27 | Never Born Protein, MW 12,3 kDa

MHSAWDLVPQLRSPHLSCINMRVIKSWDLTRMVEFLDYQNTKADGFNHKMNRATEAEIA
TDSKARGYVFVKLARTGMNSLVADLGAGKATLGHAPMKAVGLRLEHHHHHHH

31 | Never Born Protein, MW 12,4 kDa

MGGGASGFVGPQETADLNRYRYVITQRKKDNEERKSVDLYAFILDTSYGVEEKKNEYNAI
DARPGLPWNHANVGVIVVWSATNSHDPKMILGDRARARRTVLEHHHHHHH

1155 | Never Born Protein, MW 12,3 kDa

MDQTKCSSAAVSVAAHKTTNKQDADSDYFKNSTHLTPDPRVLVRNEMDTMFEVNSLILY
EHSAAALTGAQISKSMKDLCTGLEFKVYPPDMIIDAVQRHHVLLLEHHHHHHH

1259 | Never Born Protein, MW 12,5 kDa

MSLGYEASEDGTRVEAIKDGGRDGHMDIEAGGCGYWLFDHKDHRYAEHPMINTRDLDA
QMFAMFAIFYEQDTIGFPDNPWSLIILPAVSYSMMMEFAPAPTTLEHHHHHHH

1684 | Never Born Protein, MW 12,7 kDa

MFWKHLTTKQAHEKETSNAMLRHAIPTAIACGIPPQFTQKTGLVLKRSPTCNMHDPAQCT
RIKMFEVLRIWDVAMKMDYSGAKYYESYQCESTWGMPGAFLLEHHHHHHH

4954 | Never Born Protein, MW 12,1 kDa

MRVIGLSESPFDPAPKFPAASLAVGPVDGNKVPDLLLQAVLALMHTSSLKNSLNIANVVR
ANITPRRSQPMKRPSILFRHSSLERDLIIVVWNDSIKDEALEHHHHHHH

7968 | Never Born Protein, MW 12,2 kDa

MVLDMPPAAQKGPLQMAGTELAEEVKNGGQSTGHSLRSLRLDVAEGPSILDRLESHRETL
HDKMEAAHELQGKKDWSLMNQYSDRGADSEHPFLKGEVTKYLEHHHHHHH

9121 | Never Born Protein, MW 12,4 kDa

MEDPTGVPPFHHKKSHNMRKVETVNYGHKGCLVITDETNSLGVPBGVDLEGLQGAGQLAM
YKPTHGIEMARAESELIARRKEHGKMAFIIVHRELIPEEDNRFLEHHHHHHH

Svoluji k zapůjčení této práce pro studijní účely a prosím, aby byla řádně vedena evidence vypůjčovatelů.

Jméno a příjmení S adresou	Číslo OP	Datum vypůjčení	Poznámka